Deployment Performance Review of the 2019 North Pacific Observer Program

P. Ganz, C. Faunce, G. Mayhew, S. Barbeaux,
J. Cahalan, J. Gasper, S. Lowe, and R. Webster

October 2020

Deployment Performance Review of the 2019 North Pacific Observer Program

P. Ganz[1], C. Faunce[2*], G. Mayhew[3], S. Barbeaux[4],
J. Cahalan[3], J. Gasper[1], S. Lowe[4], and R. Webster[5]

1 Sustainable Fisheries Division
Alaska Regional Office
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
709 West 9th Street
Juneau, AK, 99801

2 Fisheries Monitoring and Analysis Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

3 Pacific States Marine Fisheries Commission
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

4 Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

5 International Pacific Halibut Commission
2320 West Commodore Way, Suite 300
Seattle, WA 98199

*corresponding author: craig.faunce@noaa.gov

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center

NOAA Techncial Memorandum NOAA-TM-AFSC-411

October 2020

ABSTRACT

This report contains the analyses and findings of the Alaska Fisheries Science Center's Fisheries Monitoring and Analysis Division's Fishery Monitoring Science Committee (FMSC) on the efficiency and effectiveness of observer deployment following the 2019 Annual Deployment Plan (ADP). Responses to comments by the North Pacific Fishery Management Council (Council) from the 2018 version of this report, and recommendations to improve data quality and guide the 2021 Annual Deployment Plan are also included. In 2019, there were 10 strata to evaluate: one full coverage stratum, five partial coverage observer strata defined by gear and tender designation, two partial coverage Electronic Monitoring (EM) strata defined by gear designation, one zero coverage EM research stratum, and one zero coverage stratum. Observers were deployed under trip-selection on 161 full coverage vessels that fished for 3,343 trips and 584 unique partial coverage vessel and stratum combinations (vessels can fish in more than one stratum) that fished 5,016 trips total. This was the second year in which data from the *EM HAL* stratum were used in catch accounting, and the first year in which data from the *EM POT* stratum were used in catch accounting.

Deployment in 2019 generally proceeded as planned, with few missed expectations or signs of potential bias. Realized coverage rates met expectations in 8 of 10 strata in 2019. The full coverage stratum was observed at a rate of 99.9%. The partial coverage *POT – Tender* stratum was observed at a rate of 29.5%, with a 95% confidence interval that fell above the expected coverage rate of 16.1%. The FMSC recommends that future ADPs fully integrate EM and observer deployment into one fishery monitoring program. Although EM data are now being used in catch accounting, the funding and selection of EM vessels has so far occurred separately from the processes used to optimize observer deployment in each ADP. Considering EM as a more integral part of those optimization processes will better enable analysts to avoid data gaps and take advantage of the respective abilities of each monitoring method (observers and EM). The FMSC also continues to recommend that NMFS link the Observer

Declare and Deploy System and eLandings database such that fishing trips can be uniquely identified to support the analyses presented to the Council. Such a linkage will better enable analysts to discern the contributing factors behind instances in which intended deployment is inconsistent with realized deployment.

CONTENTS

INTRODUCTION

Background of the North Pacific Groundfish and Halibut Observer Program

Fisheries observers and electronic monitoring (EM) systems collect independent information that is used to determine the effects of fishing on natural resources. The National Marine Fisheries Service (NMFS) uses its observer program in Alaska to enable the use of tools such as catch quotas to manage against the over- or under-harvest of fishes. Observers and EM are two verifiable methods for collecting fishery discard information used to estimate total catch. Observers (but not the EM systems currently used in the North Pacific) are able to record seabird and marine mammal interactions with fisheries as well. Observers also collect biological information such as length, sex, weight, ageing structures (e.g., otoliths, spines, scales, and vertebrae), and stomachs to support ecosystem studies and stock assessments.

The observer program in the North Pacific has a long history. Observers were first deployed onto fishing vessels in the Bering Sea in 1973 and into the remainder of the North Pacific in 1975 (Nelson et al. 1981, Wall et al. 1981). Fisheries in the North Pacific were initially prosecuted exclusively by foreign and later by "joint venture" operations where a developing domestic fleet of catcher vessels delivered to foreign-owned processing vessels. During the foreign and joint venture operations, foreign vessels carried fisheries observers at their expense, while domestic vessels were exempted from this observer coverage. As foreign vessels' rights to fish in the U.S. Exclusive Economic Zone (EEZ) were reduced over time and the domestic fishery grew, it became obvious to managers that observer coverage would be necessary for the emerging domestic fleet. At the onset of fully domestic fishery operations in 1990, the North Pacific Groundfish Observer Program was established as an interim observer program with rules governing observer coverage codified in regulations. This interim program would be extended four times over the next 20 years by the North Pacific Fishery Management Council (Council) - the last without a sunset date.

The regulations established in 1990 required vessels 60-125 feet in length (overall) and all vessels fishing pot gear to carry observers at their own cost for 30% of their fishing days in a calendar quarter plus at least one trip in each fishery they participate in (termed the "30% fleet"), and vessels greater than 125 feet in length to carry an observer for 100% of their fishing days at their expense. Some vessels were not required to carry observers. These included vessels less than 60 feet, vessels fishing jig gear or vessels fishing with trawl gear that deliver unsorted codends to processing vessels (termed "catcher processors" or CPs if the vessel also has catching ability and "mothership" or M if the vessel does not) and vessels that fished for Pacific halibut (*Hippoglossus stenolepis*). For shoreside processors, the rules governing observer coverage were based on the estimated tonnage processed in a calendar month: plants that processed less than 500 metric tons (t) a month were exempted from coverage, those that processed between 500 t and 1,000 t a month were required to be observed for 30% of the calendar days, and those that processed more than 1,000 t a month were required to be observed for each day in the month.

Soon after the establishment of the domestic observer program, concerns over the ability and incentive for fishers to manipulate observer coverage in a way that might bias catch estimates and other analytic products prompted efforts by NMFS and the Council to provide a mechanism for NMFS to gain control over where and when observers were deployed (Faunce and Barbeaux 2011). From 1992 to 2008, several attempts to "restructure" the program were made. In 2010, the Council unanimously decided to move forward with the restructured observer program. In 2012, the Final Rule 77 FR 70062 was published to implement Amendment 86 to the Fishery Management Plan for Groundfish of the Bering Sea and Aleutian Islands (BSAI) Management Area and Amendment 76 to the Fishery Management Plan for Groundfish of the Gulf of Alaska (GOA). Amendments 86/76 added a funding and deployment system for observer coverage to the existing North Pacific Groundfish Observer Program and amended existing observer coverage requirements for vessels and processing plants. The "restructured" North Pacific Groundfish and Halibut Observer Program (hereafter termed "Observer Program") began in 2013 with the randomization of deployments among trips and vessels. In 2018, the use of EM was added as an

additional catch monitoring tool, with the understanding that some data elements collected by observers would not be collected using EM systems.

THE ANNUAL DEPLOYMENT PLAN AND REVIEW

The restructure of the Observer Program established new annual reporting processes. Each June, the NMFS provides the Council with a comprehensive evaluation of past years' observer deployments, costs, sampling levels, and implementation issues as well as recommended changes for the coming year. This evaluation is referred to as the Observer Program Annual Report. As one chapter of the Annual Report, the deployment performance review aims to identify areas where improvements are needed to 1) collect the data necessary to manage the groundfish and halibut fisheries; 2) maintain the scientific goals of unbiased data collection; and 3) accomplish the most effective and efficient use of the funds collected through the observer fee. The annual deployment performance review is an opportunity to inform the Council and the public of how well various aspects of the program are working, and consequently lead to recommendations for improvement as appropriate. The NMFS also prepares the Observer Program Annual Deployment Plan (ADP) each fall. The ADP defines deployment strata and establishes selection rates given available budgets and anticipated fishing effort. A draft ADP is released by September of each year to allow review by the Council's Groundfish Plan Teams, as well as the Council and its Scientific and Statistical Committee (SSC). Based on input from its advisory bodies and the public, the Council may choose to clarify objectives and provide recommendations to NMFS for the ADP. Upon analysis of the Council recommendations, NMFS will make any necessary adjustments to finalize the ADP and release it to the public. The ADP is released to the public prior to the December Council meeting.

Fishery Monitoring Science Committee

Each year the Alaska Fisheries Science Center's (AFSC) Fisheries Monitoring and Analysis (FMA) Division establishes a committee to review the scientific elements of the North Pacific Observer

Program. This committee, formerly referred to as the Observer Science Committee (OSC), was renamed in 2020 as the Fishery Monitoring Science Committee (FMSC), in order to reflect the addition of EM as a tool being used to monitor fisheries in the North Pacific. Similarly, we use the term 'monitoring' in this analysis when referencing fishing activity that has been monitored either by an observer or with EM.

The FMSC provides scientific advice in the areas of regulatory management, natural science, mathematics, and statistics as they relate to deployment of fishery monitoring tools and sampling in the groundfish and Halibut fisheries of the BSAI and the GOA. The FMSC members have analytical and scientific expertise relating to fishery dependent sampling of groundfish and halibut fisheries of the BSAI and GOA and use of the collected data. If possible, the FMSC is represented by at least one member of the AFSC/FMA (Observer Program) Division, one member of the AFSC/Stock Assessment and Multispecies Assessments Program, one member of the Alaska Regional Office (AKRO) Sustainable Fisheries Division, and one member of the International Pacific Halibut Commission (IPHC).

The Sampling Design of the Observer Program

Since 2013, the Observer Program has used a stratified hierarchical sampling design with randomization at all levels. Stratification is used to increase the efficiency of sampling by observers and to address logistical issues associated with deployment. By grouping similar fishing activities into strata and sampling those strata appropriately, sampling efficiency is increased and the variance of resulting estimates may also be decreased. Sampling strata are defined in the ADP and are designed such that each unit of deployment (trip) is assigned to only one stratum.

Within a stratum, observers are deployed randomly to either vessels for a predetermined period of time (termed vessel-selection), or to individual fishing trips (termed trip-selection). In both cases, this initial deployment to the fishery is the first level of the sampling hierarchy and defines the primary sampling unit (PSU; either vessel-periods or individual trips). The list of all PSUs in a stratum defines the sampling frame and should equate to the population of interest for that sampling stratum (e.g., all trips

taken by trawl vessels fishing in the U. S. EEZ off Alaska). If the sampling frame does not contain all elements of the stratum, the resulting information may be biased. The magnitude and direction of the bias will depend on how different the fishing activities in the sample frame are from actual fishing activity.

Although this report evaluates whether monitoring goals were met, we include a brief summary of the full sampling hierarchy here for context. For each observed trip, if all hauls cannot be sampled for logistical reasons, hauls are randomly selected to be sampled. This is the next level in the hierarchy; the secondary sampling units are defined as hauls within a trip. Randomization of haul selection is designed to allow observers to record and transmit data, attend to other non-sampling responsibilities, and to allow observers time to sleep and eat. Randomization of haul selection also gives EM video reviewers the ability to optimize the amount of video that can be reviewed from each trip. Haul selection is determined using the random sampling tables and random break tables provided by NMFS. For each haul, fishing location and effort (e.g., number of hooks) are recorded, while marine mammal and seabird interactions are primarily recorded on randomly selected hauls. The ability of EM to capture marine mammal and seabird interactions is less than that of observers due to the fixed location in which EM equipment is placed.

For the randomly selected hauls, a random sample of the catch is collected (observers) or selected for video review (EM), and data from those samples are used to determine the species composition and amount of discarded catch. These samples of catch within each haul are the third level of the sampling hierarchy. While observers are trained to collect multiple large samples of catch, the number and size of samples taken from each haul will depend on the vessel configuration, fishing operations, and diversity of catch. The size of EM samples is largely determined by the number of video reviewers available relative to the amount of video to be reviewed.

At the fourth level of the sampling hierarchy, a predetermined number of individual fish of predetermined species is randomly selected from the species composition sample and measured. Lastly, at the fifth sampling level, a random selection of fish is used to collect otoliths, reproductive maturity assessments, stomach contents, genetic tissues, and other biological specimens. The number and species

of fish selected for measurement and biological specimen collection is specified each year by the AFSC's stock assessment scientists. Sampling rates for genetic tissue collection by observers (e.g., 1 of 10 Chinook salmon caught as bycatch) are set each year by the AFSC's Auke Bay Laboratories. Sampling at the fourth and fifth levels of the sampling hierarchy does not occur with EM.

More information on the sampling design used by observers and the relationship between the sample design and catch estimation can be found in Cahalan and Faunce (2020) and the 2019 Observer Sampling Manual (AFSC 2018). A summary of the 2019 ADP is included below. The focus of this report is related to deployment, and the evaluation is at the trip level of the sampling hierarchy.

THE 2019 ANNUAL DEPLOYMENT PLAN

The following briefly summarizes the final 2019 ADP (NMFS 2018). In general, all vessels that participate in cooperatives or act as catcher-processors or motherships are fully observed at the trip-level and constitute the full-coverage category of the fleet. In 2016, NMFS published new regulations to allow the owner of a trawl catcher vessel to annually request that NMFS place requesting vessels in the full coverage category for all directed fishing for groundfish using trawl gear in the Bering Sea and Aleutian Islands management area (BSAI) in the following calendar year. This regulated process has replaced an interim policy. For the 2019 calendar year, NMFS received and approved requests and placed 18 catcher vessels in the full coverage category for all directed fishing for groundfish using trawl gear in the BSAI management area (NMFS 2018). The NMFS used only the trip-selection method (i.e., no vessel-selection) to assign observers and EM to vessels in the partial-coverage category for 2019. The partial-coverage category includes vessels greater than or equal to 40 feet length overall (LOA) and not in the full coverage category. There were seven sampling strata in the partial coverage category in 2019:

1.   Hook-and-line vessels with observers (*HAL* stratum).

2.   Hook-and-line vessels with EM (*EM HAL* stratum).

3.   Pot vessels not delivering to tenders with observers (*POT - No Tender* stratum).

4. Pot vessels delivering to tenders with observers (*POT - Tender* stratum).

5. Pot vessels with EM (*EM POT* stratum).

6. Trawl vessels not delivering to tenders with observers (*TRW - No Tender* stratum).

7. Trawl vessels delivering to tenders with observers (*TRW - Tender* stratum).

Data collected through EM was first used in catch accounting in 2018, with the inclusion of *EM HAL* data. In 2019, data from the *EM POT* stratum were also used in catch accounting. Vessels had to volunteer to participate in the fixed gear EM Program by 1 November 2018. The NMFS then selected vessels to be included in the EM Program following an evaluation of available funding (NMFS 2018, Appendix C). NMFS also sought vessels to participate in EM research and development activities. In 2019, all vessels that volunteered for the fixed gear EM Program or EM research activities were selected by the NMFS and not required to carry observers, but were required to continue to log their fishing trips into the Observer Declare and Deploy System (ODDS). In this report, we attempt to evaluate the deployment of EM onto fixed gear vessels to the same degree as we evaluate the deployment of observers.

PERFORMANCE REVIEW OBJECTIVES

The following sections contain the FMSC review of the deployment of observers in 2019 relative to the intended sampling plan and goals of the 2019 ADP (NMFS 2018). This report identifies where potential mechanisms for biases exist and provides recommendations for further evaluation, including potential improvements to the observer deployment process that should be considered during the development of the 2021 ADP.

The following items from the 2019 ADP have been identified as objectives for evaluation in this report:

- Deploy observers for the planned number of sea days. This objective will be considered to be met if the actual number of sea days expended falls within the range of values from simulated sampling provided in the 2019 ADP. The Observer Program's budget was expected to cover 3,110 days in 2019.

- Deploy observers and EM at the coverage rates specified in the 2019 ADP. Following the 2019 ADP, ODDS was programmed to randomly select logged trips at a rate of 23.70% in the *TRW - No Tender* stratum, 17.71% in the *HAL* stratum, 15.43% in the *POT - No Tender* stratum, 27.12% in the *TRW - Tender* stratum, 16.11% in the *POT - Tender* stratum, and 30% in the EM strata. Under a randomized deployment scheme, these partial coverage selection rates are expected to be within a 95% confidence interval computed from the realized coverage rates.

- Collect tissue samples from Chinook and chum salmon bycatch as specified in the 2019 Observer Sampling Manual to support genetic analysis and identify stock of origin.

- Collect observer and EM samples that are representative of the entire fishing fleet (observed and monitored trips are equivalent to unobserved and unmonitored trips within a stratum). Evaluation of this objective is focused on the randomization of observer and EM deployments into primary sampling units, and how departures from a random sample affect data quality.

Observer Deployment Performance Metrics

Performance metrics have been developed to assess whether the trip-selection process (through the implementation of the 2019 ADP) provides a representative sample of fishing trips in the North Pacific in 2019. These metrics reflect four mechanisms that can impact the quality of the data: sample frame discrepancies, non-response, differences in trip characteristics, and sample size.

The performance metrics used in this evaluation are as follows:

1. Deployment rates for each stratum: This is the basic level of evaluation for comparing targeted and achieved sampling rates, where sampling strata are partitions of the entire population about which we want to make inferences (e.g., generate estimates of catch). Implementation challenges can be identified in this step, such as sample frame inadequacy, selection biases, and issues with sample unit definitions. Specifically, this section assesses the following:

    a. Sample rates and number of samples relative to intended values.

    b. Quantification of under- and over-coverage rates (sample frame discrepancies). Over-coverage of a population occurs when the sample frame includes elements that are not part of the target population. When these elements are included in the random sample, effort (time, cost) is expended needlessly. Under-coverage results from having a sample frame that does not include a portion of the target population which can lead to biased data if that portion of the population differs from the population included in the sample frame.

    c. Non-response rates. Non-response occurs when randomly selected elements (trips or vessels) are not actually sampled. If these trips or vessels have different fishing behavior (e.g., catch, areas fished) than the rest of the population, the data collected will not represent the entire fleet (non-response bias).

2. Representativeness of the sample: Randomized sampling is a method used to ensure that the results of sampling reflect the underlying population. Departures from randomization can lead to non-representative data and hence potential bias in estimates of the parameters of interest. A randomized sample design is expected to achieve a rate of monitored events that is similar across both space and time. Representativeness of the sample was divided into three separate components:

a. Temporal representativeness

   i. Effort plots: plots of expected and actual monitoring effort over time. Areas where these two lines deviate from each other are indicative of periods with differential realized sample rates (and potential temporal bias).

b. Spatial representativeness

   i. Maps: Maps provide a visual depiction of the spatial distribution of monitoring coverage relative to effort in each partial coverage stratum, as well as where low or high coverage rates occurred.

   ii. Probability of monitoring a fewer or greater number of trips within an area than would be expected given the realized sample rate for the entire stratum. These data are used to identify departures from anticipated sampling rates.

c. Representativeness of trip characteristics

   i. Consistency of trip characteristics for monitored and unmonitored portions of the stratum. These metrics are based, in part, on the availability of data for both monitored and unmonitored fishing activities; for example, data that are reported for all trips on landing reports. Attributes tested in this report include the following:

       - Trip duration (days).

       - Vessel length (feet).

       - The number of NMFS Areas visited during the trip.

       - The amount of landed catch (metric tons).

       - The number of species in the landed catch (also known as species richness).

- The proportion of the total landed catch that was due to the most prevalent species (pMax, an inverse measure of species diversity where an increase in pMax indicates a decline in diversity).

3. Adequacy of sample size: A well-designed sampling program will have a sample large enough to reasonably ensure that the characteristics of interest in the entire target population are represented in the data. Whether the sample size collected was adequate was determined through an examination of the probability of deploying observers at the implemented rate and having no monitoring coverage in one or more cells (e.g., defined by NMFS Reporting Area and strata).

Although these metrics can identify places where observed results differ from expectations, it is ultimately a subjective decision as to whether or not these differences are substantial enough to have management implications. This holds true even for tests that have associated *p*-values. Additionally, our focus on landed catch is due to the fact that total catch is comprised of retained and discarded portions, and since discarded catch is not available from unmonitored trips, landed catch represents the only portion of the catch that is available from all trips.

CHANGES TO THIS REPORT FROM LAST YEAR

This year we made several updates to our analyses. These include two major and several minor changes. The first major change this year is the addition of a new analysis of data gaps (Appendix B). Following the methods used in the gap analysis in Appendix C of the Draft 2020 Annual Deployment Plan (NMFS 2019a), Appendix B serves to evaluate the extent to which monitoring coverage within deployment strata was distributed proportionately to post-strata defined by Fishery Management Plan Area (FMP) and trip target (predominant species) by evaluating the spatiotemporal proximity of monitored trips to unmonitored trips and assesses the likelihood of acquiring the achieved coverage in 2019 given the assumption of random deployment. It is the intent that elements of this Appendix be

included in future Annual Reports. The second major change is how we calculate *p*-values in the permutation tests that assess whether or not observed or unobserved trips were different in a given metric. In recognition that these tests are not independent within a stratum, this year we adjust *p*-values to account for multiple comparisons by multiplying by the number of tests performed. Thus we inflate each *p*-value to reduce the chances of making a false interpretation of differences where there are none. This is known as a Bonferroni adjustment, and is also applied to permutation test *p*-values in Appendix A. Results now focus on only large differences. Minor changes to the tables in this chapter include the addition of a table showing the average review times for fixed gear EM video, the realized cost of the partial coverage monitoring program in dollars with the expected cost of the program, and more information on when trips in ODDS were selected due to the cancellation of prior trips.

EVALUATION OF DEPLOYMENT IN 2019

The deployment of observers into the 2019 Federal fisheries in Alaska is primarily evaluated at the level of the deployment stratum because each stratum is defined by a different sampling rate or by a different monitoring method (e.g., observers and EM). In this document, trips in the *EM HAL* and *EM POT* strata are considered successfully monitored if at least some video was reviewed from a trip. The rationale for defining monitored trips this way is that it is most similar to the way in which trips in other strata are considered observed (i.e., irrespective of whether or not haul information or usable species composition data were collected).

Evaluating Effort Predictions

Each year, the NMFS sets an annual budget for the Observer Program in terms of cost and observer days. The partial coverage observer day budget for 2019 was set at $4,452,623 and 3,110 days in the 2019 ADP, and the NMFS expected to spend $4,342,097 observing 3,109 days (NMFS 2018). The expected number of observer days is determined by the expected number of fishing days and the rate at

which trips are selected for coverage. The number of fishing days expected to occur in 2019 was estimated using data on annual fishing effort from 2013 to 2018 (Ganz and Faunce 2019). Based on simulations using trip durations from 2017 and 2018, the NMFS then set selection rates so that the average cost from simulations was equal to the available budget (NMFS 2018).

In 2019, the FMA paid for 3,315.5 observer days, which was 6.6 % greater than predicted by the average simulation, but well within the range of possibilities predicted in the 2019 ADP (Fig. 1, top panel). This is explained by the fact that there was more effort in *HAL*, *POT – Tender*, and *TRW – Tender* than expected (Table 1). Despite observing more days than predicted, expenditures for partial observer coverage were under budget (Fig. 1, bottom panel). This resulted because the cost of a partial coverage observer day in 2019 was less than the expected cost that was estimated in the 2019 ADP.

PERFORMANCE OF THE OBSERVER DECLARE AND DEPLOY SYSTEM
IN TRIP-SELECTION

The random selection of trips for monitoring is made by the ODDS for logged trips within the observer and EM trip selection pools. The ODDS generates a random number according to the pre-determined rates and assigns each logged trip to either "selected to be monitored" (selected) or "not selected to be monitored" (not selected) categories. The NMFS monitoring providers have access to all selected trip information necessary to facilitate logistics. Up to three trips may be logged in advance of fishing to provide industry users with flexibility to accommodate their fishing operations.

Logged trips have different dispositions. When initially logged, trips are considered pending and can be either closed or cancelled. Whether changes can be made by the user (person logging the trip) or must be made by the monitoring provider (or the NMFS) depends on whether or not the trip is selected to be monitored, the stratum the trip belongs to, and the timing of the activity. Trips can be closed (marked as complete) by the ODDS user after the planned trip departure date by either entering the dates of the trip

and the port processor of the landing, or by selecting from a list of pre-populated landing reports. For partial coverage strata monitored by observers, the observer provider is given 72 hours prior to the trip start to provide for an observer to board the vessel. While a trip may be entered into ODDS that is scheduled to start earlier than 72 hours from the time of entry, if selected for observer coverage, the observer provider can opt to delay the start of the trip up to, but not exceeding 72 hours from the time of trip entry. This helps protect the observer provider from the high cost of deploying an observer with short notice. The vessel operator is protected as well by guaranteeing the assigned observer to the vessel up to 48 hours past the planned start of the fishing trip. This rule helps ensure that an observer is available to the boat in case of unforeseen events such as weather. If, however, the trip start date and time has passed by more than 48 hours, then the observer provider can cancel the trip and release the observer from the vessel and trip, and the vessel would need to log a new trip with a new 72-hour notice in place prior to fishing. These 'forced cancellations' are not present in trips that are not selected for observation since the logging, closing, or cancellation of the trip is entirely under vessel control. The vessel operator may change the dates of a logged trip regardless of selection status prior to, or in lieu of cancellation. However, trips that have not been closed at the end of the calendar year are automatically cancelled by the ODDS to prevent past trips from affecting the deployment rates set for the next year's ADP.

The number of trips logged in the ODDS in 2019 and their dispositions is summarized in Table 2, Table 3, and Table 4. The cancellation rate by users and by the ODDS is summarized for selected trips in each stratum (Table 2). Of the 5,513 total trips logged, 1,264 were selected, and 226 were cancelled: 3 by ODDS (0.24%) and 223 by users (17.6%). The user cancellation rate for selected trips ranged from 1.9% for *EM POT* to 26.7% for *TRW - Tender*.

The flexibility offered by the ODDS means that the outcome of random selection is known to the vessel operator for up to three logged trips in advance of fishing. In the case where ODDS users disproportionately cancel selected trips, one would expect monitoring coverage to be lower than the programmed selection rates. To reduce this potential bias, the ODDS is programmed to automatically select the vessel's next logged trip if a previously selected trip was cancelled by the user. Although these

"inherited" trips preserve the *number* of selected trips in the year, they cannot prevent the *delay* of selected trips during the year. Therefore, the potential for temporal bias is still present. The percentages of selected trips from either inherits or waivers are found in Table 3. The relative percentage of selected trips that inherited their final selected-status due to a previous cancellation ranged from 3.8% for *EM POT* to 26.7% for *POT - Tender* (Table 3). Within the same gear-type, cancellation rates and the proportion of inherited trips were much larger for strata that used observers for at-sea monitoring than those that used EM.

The extent to which trip-selections are changed from the time they are entered can be determined by comparing the rate of trip observation expected from 1) random selection of all logged trips (initial random selection) and 2) random selection of remaining trips after cancellations, waivers, and inherited trips. In any case, the proportion of trips selected to be observed should fall within what would be expected given the binomial distribution (since each trip is either selected or not selected). The rates obtained (%, with associated *p*-value based on the binomial distribution) in the initial selection process were within expected ranges with the following exceptions – the initial selection rate was 33.91% (*p*-value = 0.011) for the *EM HAL* stratum, and 39.47% (*p*-value = 0.020) for the *TRW - Tender* stratum (Table 4). This means that the *EM HAL* and *TRW - Tender* strata were being over-selected in ODDS, and that we should interpret high final coverage rates in these strata with caution.

The final selection rate after trips were closed, cancelled, or waived were within expected bounds with the exception of the *HAL* stratum 20.47% (*p*-value = 0.006), the *EM HAL* stratum, 34.80% (*p*-value = 0.002) and the *TRW - Tender* stratum 46.55% (*p*-value = 0.002; Table 4). Given the high initial selection rates, we can safely disregard these final selection rates with the exception of the *HAL* stratum.

Differences in the initial selection rates of ODDS and those that result after cancellation and trip changes can also be looked at over time (Fig. 2). In this plot, we are mostly concerned when the lines representing the two selection rates in this plot diverge substantially. This pattern can occur when cancelled trips that were originally selected for coverage are preserved through the inherit process, while cancelled trips that were not originally selected for coverage are not.

15

In addition to the inherit process, the lack of linkage between the ODDS and eLandings contributes to the differences between programmed selection rates in ODDS and trips that are ultimately observed. Currently, ODDS provides users with a list of Report IDs from eLandings from which to close their logged trips. However, these data are not validated or error checked, making them unreliable in their current state. This linkage between the logged (ODDS) trip (with its selection probability) and its associated landing information is necessary to evaluate potential improvements in deployment efficiency within the partial coverage fleet.

Evaluation of Deployment Rates

This section compares the coverage rate achieved against the expected coverage rates. Data used in this evaluation are stored within the Catch Accounting System (CAS, managed by the AKRO), the Observer Program database (NORPAC, managed by the AFSC), and eLandings (under joint management by Alaska Department of Fish and Game - ADF&G; the International Pacific Halibut Commission - IPHC; and the NMFS). Separate rate evaluations are conducted depending on whether the unit of observer deployment was at-sea fishing trips or dockside deliveries of pollock.

At-sea Deployments

The 2019 Observer Program had 10 different deployment strata to be evaluated (Table 5). There was one full coverage stratum comprised of trips taken both by vessels that were required to have full coverage (e.g., AFA vessels) and those fishing in the BSAI that opted into full coverage. There were seven partial coverage strata: five observed strata defined by gear and tender designation and two EM strata defined by gear designation. There were also two zero coverage strata: one zero coverage EM research stratum and one zero coverage stratum for jig vessels and vessels under 40 ft. length overall.

Evaluations for the full coverage category and zero-selection pool are straightforward - either the coverage achieved was equal to 100% or 0%, respectively, or it was not. The program achieved 99.9% coverage in its full coverage category (Table 5). Five trips were not monitored in the full coverage

category – four of these occurred on a single catcher vessel fishing CDQ halibut with hook-and-line gear that retained Pacific cod above the Maximum Retainable Amount and therefore met the criteria for full coverage fishing but failed to obtain a full coverage observer. The program achieved perfect compliance with the zero coverage stratum (Table 5). Under the assumption that the deployment was randomized, a 95% confidence interval computed from the realized coverage rates (under the assumption of a binomial distribution for observed trips) will contain the actual deployment rate 95% percent of the time. If expected coverage levels (ODDS programmed rates) were within the 95% confidence intervals, then we conclude that realized and expected coverage rates were equal. Coverage rates were consistent with expected values in six of the seven partial coverage strata, but were higher than expected within the *POT - Tender* stratum (Table 5). There are two reasons why this result is of little concern. First, there is no clear evidence of trip manipulation in ODDS data from this stratum. Secondly, the expected rate was only slightly outside of the 95% confidence interval that surrounds the realized rate (16.1% expected vs. a confidence interval of 16.8% to 45.2%). Given the low number of total trips in this stratum (44), a change in a single observed trip, from 13 observed to 12 would have resulted in an expected result for this stratum since new confidence bounds would have included the expected rate.

Unlike observed trips, the coverage rate for EM is based on information provided from the Pacific States Marine Fisheries Commission (PSMFC) that is available to analysts in the NORPAC database. By the end of 2019, the PSMFC had reviewed nearly all of the EM hard drives received (Table 6). In 2019, the mean time between receipt and completion of review was 58 days for *EM HAL* and 79 days for *EM POT* (Table 7). This is compared to an average of 8.8 days during pre-implementation in 2016 (NMFS 2017, p. 87).

In combination across all strata, coverage levels, and fishery monitoring tools, 4,497 trips (43.3%) and 510 vessels (47.0%) were successfully monitored at-sea among all fishing in Federal fisheries of Alaska in 2019 (Table 5).

Observers were assigned to monitor shoreside deliveries of pollock. The objective of this monitoring was to obtain a count of the number of salmon caught as bycatch and to obtain tissue samples for genetic analysis from these fish in each observed pollock delivery. The sampling design used for this objective in 2019 remained unchanged from that used since 2011 (Faunce 2015); all deliveries of pollock that were observed at sea were also observed dockside. While all Bering Sea pollock trips and deliveries are observed, this is not the case in the GOA, where pollock trips randomly selected for at-sea monitoring are also expected to be sampled shoreside for salmon (NMFS 2013). For this analysis, pollock deliveries are defined as any delivery where the predominant species is pollock in eLandings.

Given the design, the level of dockside observation of walleye pollock deliveries should be 100% in the full coverage category. In 2019, 100% of full coverage walleye pollock deliveries were observed (Table 8).

While dockside monitoring expectations of the full coverage category are straightforward, evaluations of the partial coverage category are more complex. For example, in the partial coverage trawl pollock fishery, non-tendered trips selected for at-sea monitoring are expected to be monitored dockside for salmon, while tendered trips are excluded from dockside monitoring as a matter of policy. While it may seem intuitive that the expected coverage rate for GOA pollock deliveries within the *TRW - No Tender* stratum should be equal to the programmed trip selection rate of 23.70%, this assumption is likely untrue because observers are not deployed into the pollock fishery but into the entire trawl fishery, and the relationship between the number of deliveries and trips is not expected to be constant, especially when measured across ports. Therefore, we present the dockside observation rates for *TRW - No Tender* pollock landings but make no comparison to deployment rates (Table 8).

Bycatch estimates of Chinook salmon in the GOA are estimated using methods described in Cahalan et al. (2014). In the event that a delivery cannot be monitored (e.g., the case in a tendered delivery or non-pollock delivery), then estimation of bycatch comes by applying salmon bycatch rates to

landed catch. Estimates of stock of origin from salmon bycatch are produced by the AFSC's Auke Bay Laboratories (e.g., Guthrie et al. 2019, Guthrie et al. 2020).

SAMPLE QUALITY

Temporal Patterns in Trip-Selection

The cumulative number of fishing trips in each stratum was multiplied by the stratum-specific selection rate to obtain the expected number of observed trips. Under the assumption that there is no temporal bias in observer coverage, 2.5% of values should be larger than the upper 95% confidence limit and 2.5% should be smaller than the lower limit. At the end of 2019 the number of observed trips was outside of this expected range in only one of the seven partial coverage strata: *POT - Tender* (expected rate = 0.161, realized rate = 0.295, *p*-value = 0.023; Table 5 and Fig. 3). Coverage rates were outside of the expected range for 15.9%, 9.3%, 31.2%, 28.2%, and 7.9% of the year for the *EM HAL*, *EM POT*, *POT – No Tender*, *POT - Tender* and *TRW – No Tender* strata, respectively. The *EM HAL*, *POT – No Tender*, and *TRW – No Tender* strata were outside of the expected range earlier in the year but fell within the expected range by the end of April (Fig. 3). Coverage rates were within their expected ranges for 100% of the year for the *HAL* (expected rate = 0.177, realized rate = 0.176, *p*-value = 0.925) and *TRW - Tender* (expected rate = 0.271, realized rate = 0.357, *p*-value = 0.175) strata. Overall, there appeared to be less temporal bias in 2019 than in 2018, when three of six partial coverage strata had coverage rates outside of the expected range at the end of the year (AFSC and AKRO 2019).

Spatial Patterns in Trip-Selection

Under a strictly random selection of trips and with a large enough sample size, the spatial distribution of monitored trips should reflect the spatial distribution of all trips. The hypergeometric distribution was used to describe the results of sampling from a population of items (fishing trips) with different characteristics (NMFS Area fished). Based on this distribution, the expected number of

monitored trips in a stratum and area is the realized monitoring rate (not selection rate) for the stratum multiplied by the total number of trips from that stratum that occurred in the area of interest. Using this method, we compared the expected number of monitored trips to the realized number of monitored trips in each NMFS Area and stratum combination and found that in most cases, the realized number of monitored trips was close to the expected result (Fig. 4). As part of this evaluation, we calculated the probability of monitoring the realized number of monitored trips within each stratum and NMFS Area. For the purposes of the following discussion, NMFS Areas with an unexpected number of trips (probability of our result is less than 0.05) are referred to as "low-p" areas.

The EM HAL stratum

Given that there were 16 NMFS Areas fished in *EM HAL*, we would expect there to be $0.05 \times 16 = 1$ low-p area for this stratum. There was one. The percent of trips monitored among NMFS Areas in this stratum ranged from 0% to 66.7% (median = 29.9%). The probability of these monitoring rates or rates that deviated further from expected values is depicted in Figure 5. These results mean that there was no clustering of monitored trips among NMFS Areas that was different from expected. No spatial bias appears to have occurred in the *EM HAL* stratum.

The EM POT stratum

Given that there were 11 NMFS Areas fished in *EM POT*, we would expect there to be $0.05 \times 11 = 1$ low-p area for this stratum. There was one. The percent of trips monitored among NMFS Areas in this stratum ranged from 0% to 80% (median = 36.8%). The probability of these monitoring rates or rates that deviated further from expected values is depicted in Figure 6. These results mean that there was no clustering of monitored trips among NMFS Areas that was different from expected. No spatial bias appears to have occurred in the *EM POT* stratum.

The HAL stratum

Given that there were 18 NMFS Areas fished in *HAL*, we would expect there to be $0.05 \times 18 = 1$ low-p area for this stratum. There were two. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 28.6% (median = 17.4%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 7. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *HAL* stratum.

The POT - No Tender stratum

Given that there were 14 NMFS Areas fished in *POT - No Tender*, we would expect there to be $0.05 \times 14 = 1$ low-p area for this stratum. There was one. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 37.5% (median = 12.2%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 8. These results mean that there was no clustering of observed trips among NMFS Areas that was different from expected. No spatial bias appears to have occurred in the *POT - No Tender* stratum.

The TRW - No Tender stratum

Given that there were 9 NMFS Areas fished in *TRW - No Tender*, we would expect there to be $0.05 \times 9 = 0$ low-p areas for this stratum. There was one. The percent of trips observed among NMFS Areas in this stratum ranged from 14.3% to 50% (median = 22.3%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 9. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *TRW - No Tender* stratum.

The POT - Tender stratum

Given that there were 7 NMFS Areas fished in *POT - Tender*, we would expect there to be 0.05 × 7 = 0 low-p areas for this stratum. There were two. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 100% (median = 16.7%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 10. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *POT - Tender* stratum.

The TRW - Tender stratum

Given that there were 5 NMFS Areas fished in *TRW - Tender*, we would expect there to be 0.05 × 5 = 0 low-p areas for this stratum. There were two. The percent of trips observed among NMFS Areas in this stratum ranged from 20% to 75% (median = 36.4%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 11. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *TRW - Tender* stratum.

Trip Metrics

This section analyses whether monitored trips are similar to unmonitored trips using a permutation test (a.k.a., randomization test). This test evaluates the question "How likely is the difference we found if these two groups have the same distribution (in the metric we are comparing)?" Permutation tests compare the actual difference found between two groups to the distribution of many differences derived by randomizing the labels defining the two groups (e.g., monitored and unmonitored). Difference values in the permutation test were calculated by subtracting the mean metric value for the "No" condition from the mean metric value for the "Yes" condition. For example, the difference between vessel lengths in a permutation test for a monitoring effect would be the mean value for unmonitored trips subtracted from the mean value for monitored trips. By randomizing group assignments, the combined

distribution of randomized differences represents the sampling distribution under the null hypothesis that the two groups are equal. In this report, 1,000 randomized trials were run for the permutation test. The p-value from the test is calculated as the number of randomized trials with greater absolute differences than the actual difference divided by the number of randomized trials. Similar to the other statistical tests used in this report, low p-values (< 0.05) indicate unlikely events under the hypothesis of equality and are therefore considered evidence against that hypothesis. As stated previously, a Bonferroni adjustment has been applied to these p-values by multiplying original p-values by the number of metrics being tested (six in this case). These adjusted p-values are then compared to the 0.05 significance level. In an attempt to improve clarity, although five values are calculated in the test; 1) the difference between groups, 2) the mean difference between groups from randomized trials, 3) #1 expressed as a percentage of the mean value of the metric being tested, 4) #2 expressed as a percentage of the mean value of the metric being tested, and 5) the p-value of the test, only values (1), (3) and (5) are presented.

Six trip metrics were examined in the permutation test. These metrics were as follows: the number of NMFS Areas visited in a trip, trip duration (days), the weight of the landed catch (t), the vessel length (ft), the number of species in the landed catch, and the proportion (0 to 1) of the total catch that is made up of the most predominant species (pMax). The metric 'vessel length' is used to help interpret the results from 'weight of landed catch' since fishing power is positively correlated to vessel length. Specifically, differences in weight *and* length are interpreted as a failure to achieve a random sample of vessels of different sizes, whereas differences in weight only lend more evidence that there was a monitoring effect. The number of species within the landed portion of the catch is a measure of species richness. Our pMax metric follows the concepts behind Hill's diversity number N1 that depicts the number of abundant species (Hill 1973) and is a measure of how "pure" catch is since a value of one would indicate that only the predominant (and presumed desirable) species was landed.

Were monitored trips similar to unmonitored trips?

The sample sizes available to the permutation test are presented in Table 9. Results of permutation tests are presented in Table 10. A visual depiction of individual results of this permutation test for the *HAL*, *POT - No Tender*, and *TRW - No Tender* strata is given in Figure 12 for illustration purposes.

- Of the six metrics compared in the *EM HAL* stratum, one had a low *p*-value. On average, monitored trips in this stratum landed 13.4% (0.52) more species than unmonitored trips. Landed catch on monitored trips was 3.5% more diverse than unmonitored trips, although this difference was borderline to the traditional 0.05 significance cutoff.

- Of the six metrics compared in the *EM POT* stratum, one had a low *p*-value. On average, monitored trips in this stratum landed 21.2% (0.49) more species than unmonitored trips.

- Of the six metrics compared in the *HAL* stratum, two had low *p*-values. On average, observed trips in this stratum were 12.3% (0.66 days) shorter in duration and landed catch that weighed 13.6% (0.90 metric tons) less than unobserved trips.

- Of the six metrics compared in the *POT - No Tender* stratum, none had low *p*-values.

- Of the six metrics compared in the *POT - Tender* stratum, one had a low *p*-value. On average, observed trips in this stratum landed catch that weighed 100.1% (175.76 metric tons) more than unobserved trips.

- Of the six metrics compared in the *TRW - No Tender* stratum, two had low *p*-values. On average, observed trips in this stratum occurred in 4.4% (0.05) fewer areas and landed 11.9% (0.73) fewer species than unobserved trips.

- Of the six metrics compared in the *TRW - Tender* stratum, there were no metrics with low *p*-values.

Based on these combined results, differences between monitored and unmonitored trips were found for species richness, trip duration, areas fished, and landed catch (Table 10). Monitored EM trips of both hook and line and pot gear types resulted in greater species numbers reported in the landings data than

unmonitored trips. If monitored and unmonitored trips occur in the same fisheries, it is possible that species are lacking on unmonitored trips or are being incorrectly accounted for on monitored trips, or that there is more at-sea discard of species on unmonitored trips. The *HAL*, *POT - Tender*, and *TRW – No Tender* strata also exhibited observer effects, although the magnitude of differences for *TRW – No Tender* was small. Of these, the *POT – Tender* result is the most striking due to the large magnitude of difference, but also the easiest to explain. Landings of tendered trips can be quite large on rare occasions, and when rare large landings occur, whether they are observed or unobserved, these single trips can 'tip the scales' for permutation tests across the entire strata. In 2019, one of these very large-landing trips was observed. However, we cannot dismiss the possibility that we incorrectly accounted for the linkages between landings and trips, and some tendered trips were actually larger, or smaller, than we calculated. More on this topic is discussed in our recommendations section.

Gear, tender, and observed status combinations

One of the analyses done by the permutation test is to compare trip lengths (in days) between monitored and unmonitored trips and determine whether there were significant differences. However, these permutation tests do not visually map the data for monitored and tendered states together. To accomplish this, a plot of the trip durations for these states is included as Figure 13. These plots illustrate *HAL* non-tendered trips were shorter in duration when observed, which was also seen in permutation tests. In addition, tendered *POT* and *TRW* trips of more than ten days appear to have been observed at a greater frequency than unobserved trips. If these longer trips also were associated with greater landed weight, then this would explain the permutation results for these strata that showed greater landed weights on observed trips compared to unobserved trips.

ADEQUACY OF THE SAMPLE SIZE

In a well-designed sampling program, the monitoring rate should be large enough to reasonably ensure that the range of fishing activities and characteristics are represented in the sample data. The Catch Accounting System post-stratifies data into groups of fishing activities with similar trip characteristics such as gear, trip targets, and NMFS Area (Cahalan et al. 2014). At low numbers of trips and low sampling rates, the probability of no monitoring data within a particular post-stratum is increased and may result in expansions of bycatch rates from one type of fishing activity against landings for a different type of fishing activity. This will result in biased estimates of bycatch. For this reason, it is important to have a large enough sample (monitored trips and vessels) to have reasonable expectation of monitoring all types of fishing.

Over the course of an entire year, some NMFS Areas have low fishing effort and as a result have a relatively high probability of being missed by the simple random sampling represented by observer deployments and EM. The fishing effort data for each stratum and the number of monitored trips over the course of 2019 were used to illustrate their combined effect on the probability of a NMFS Area containing monitoring data using the hypergeometric distribution (Fig. 14). From this figure it can be seen how 1) the likelihood of at least one monitored trip is increased with fishing effort and 2) is also increased with an increase in the selection rate. Given our sampling rates in the 7 partial coverage trip-selection strata, the probability of having no monitored trips in a NMFS Reporting Areas increases quickly above 0.05 when there are fewer than 8 trips in the *EM HAL* stratum, 6 trips in the *EM POT* stratum, 15 trips in the *HAL* stratum, 19 trips in the *POT - No Tender* stratum, 7 trips in the *POT - Tender* stratum, 10 trips in the *TRW - No Tender* stratum, and 6 trips in the *TRW - Tender* stratum in a given area. Including additional factors such as week, gear, and target will decrease the number of trips with the same characteristics and hence increase the probabilities of obtaining no monitoring data of that character (post-strata of the CAS).

A new analysis presented in Appendix B – Gap Analysis examines the deployment of observers and EM systems at finer spatiotemporal scales than presented here. This new analysis evaluates the availability of monitoring coverage within and between the partial coverage selection pools and highlights instances where sampling effort was disproportionately distributed in space and time between post-strata defined by gear, NMFS Area, and dominant species landed (trip target). For example, the spatial patterns in the *HAL* stratum appear to be due to disproportionately high monitoring rates in the GOA for trips targeting halibut and lower monitoring rates for halibut-target trips in the BSAI, especially in the Aleutian Island areas. Additionally, the low number of observed trips in area 620 in the *TRW – No Tender* stratum was due to disproportionately low monitoring rates among arrowtooth-target trips where only 2 of 42 trips were observed (Appendix B – Gap Analysis).

RESPONSE TO COUNCIL AND SSC COMMENTS

The SSC has requested that a specific section with responses to SSC comments be provided in the written report, as is done for SAFE documents. This section addresses FMSC responses (in italics) to comments relative to this chapter made by the Council and the SSC after the presentation of the 2018 Annual Report during the June 2019 Council meeting.

Council comments:

In the 2019 Annual Report (to be presented in June, 2020), the Council recommends that NMFS:

- Continue to include an evaluation of observer effects in pelagic and non-pelagic trawl within the trawl stratum.

  This evaluation is included as Appendix A.

SSC recommendations:

The SSC offered the following recommendations to the Observer Program:

- The analysts to initiate a comparison of the likely magnitude of bias that has been detected between monitored and unmonitored trips with the overall magnitude and precision of discard or PSC that is being monitored for compliance by management.

  While some differences were detected between monitored and unmonitored trips, the impact that these types of differences have on estimates of discard is not known at this time. We note that in 2019, the detected differences occurred primarily within the hook and line stratum where fishing activity is not limited by PSC or bycatch quotas.

- Consider [EM] coverage for the under-40'-no coverage fleet for 2019.

  This was not considered in the 2020 ADP.

- In cases where there are multiple gear types in a stratum (e.g., pelagic and non-pelagic trawls) the SSC recommends analysis of the results by gear type separately in addition to analysis aggregated to the stratum level. Such disaggregation will avoid masking of gear-specific differences in catch composition and other factors that could provide justification for possible further subdivision of strata.

  We included an evaluation of observer effect tests for different types of trawl gear in Appendix A. In response to the SSC recommendation, we note that further subdivision of strata may not be feasible as total sample size continues to decline. For example, from 2019 to 2020 the ability of observer data to adequately sample tendered and non-tendered strata was compromised to the point that the designation was no longer supported, and these trip types are now combined for a gear-based stratum.

- The SSC looks forward to seeing a full evaluation of this [EM] program as soon as is practical, as well as an evaluation of the tradeoffs between use of EM and the existing partially observed coverage category. As the Council considers continued growth of the EM program, it will be important to conduct appropriate cost comparisons, specifically including video review costs,

as well as an evaluation of the ability of EM versus onboard observer data to meet program needs.

Costs are not addressed in this analysis.

FMSC RECOMMENDATIONS TO IMPROVE DATA QUALITY

Recommendations from the 2018 Annual Deployment Review

The Fisheries Monitoring Science Committee (formerly the Observer Science Committee) made the following recommendations in its 2018 review of observer deployment to be considered in developing the 2020 ADP (NMFS 2019b). Following each recommendation is the italicized outcome of that recommendation.

The Fishery Monitoring Science Committee's Recommendations to improve the 2020 ADP were as follows:

- The ODDS trip logging and cancellation rules be re-evaluated and communicated to the Council and industry as soon as possible.
  No formal public action has been taken by the NMFS.
- The draft 2020 ADP stratification designs include a re-examination of tendering strata.
  The distinction between tendered and non-tendered strata was eliminated in the 2020 ADP.
- Do not stratify by type of trawl gear (i.e., NPT and PTR strata).
  These gear types were not separated in the 2020 ADP. The rationale for not creating separate NPT and PTR strata is included in Appendix A.
- Continue the baseline + optimization approach for determining coverage levels among strata.
  The 15% baseline + optimization approach for determining coverage levels among strata for observer coverage was used in the 2020 ADP.

- We recommend that EM review rates be set to ensure that the entire year is sampled and review is timely enough so that data from EM can be used for catch accounting and fisheries monitoring as envisioned by the Council.

   EM review included the entire year for 2019, which was an improvement over 2018. There was about a two-month lag between data collection and data availability in 2019. Whether this is timely enough for catch accounting and fishery monitoring is not clear to the FMSC.

Recommendations to Improve Data Quality and Guide the 2021 ADP

1. **We recommend that the ADP fully integrate EM and observer deployment into one fishery monitoring program.** This recommendation echoes the SSC recommendation made at their June 2019 meeting, and is based on the recognition that EM and observers are two tools at the disposal of the NMFS to monitor fisheries and each has its advantages and disadvantages. Issues due to incomplete integration of fishery monitoring tools occurred in 2019 when only EM trips were monitored in the pot gear Pacific cod Central Gulf (Area 630) fishery, introducing a data gap for the GOA Pacific cod stock assessment. In 2020, observer coverage has been reduced further as a result of COVID-19 precautions.

2. **We continue to recommend that NMFS link the ODDS and eLandings database such that fishing trips can be uniquely identified to support the analyses presented to the Council.** The analyses contained here attempt to identify fishing trips, which is the unit of measurement for deployment. However, there are some instances when realized deployments do not match intended deployments. In some cases, it may be that there were no differences, but the accounting of trips between ODDS and eLandings data are incongruent. We note that the temporal bias issue identified (Figure 3) in the observed tendered pot stratum and differences between the observed and unobserved landed weight (Table 10) in this stratum was potentially an artifact of the analysis. This artifact could have

been caused by the difficulty in identifying observed and unobserved trips, especially for

tendered strata.

CITATIONS

AFSC. 2018. 2019 Observer Sampling Manual. Fisheries Monitoring and Analysis Division, North

    Pacific Groundfish Observer Program. AFSC, 7600 Sand Point Way N.E., Seattle, Washington,

    98115.

Cahalan, J., J. Mondragon, and J. Gasper. 2014. Catch sampling and estimation in the Federal groundfish

    fisheries off Alaska: 2015 Edition. NOAA Tech. Memo. NMFS-AFSC-286, 46 p. Available at

    https://repository.library.noaa.gov/view/noaa/4833.

Cahalan, J. and C. H. Faunce. 2020. Development and implementation of a fully randomized sampling

    design for a fishery monitoring program. Fishery Bulletin (U.S.) 118:87-99.

Faunce, C. H., and Barbeaux, S. J. 2011. The frequency and quantity of Alaskan groundfish catcher-

    vessel landings made with and without an observer. ICES J. Mar. Sci. 68:1757-1763.

Faunce, C. H. 2015. Evolution of observer methods to obtain genetic material from Chinook salmon

    bycatch in the Alaska pollock fishery. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-

    288, 28 p. doi.org/10.7289/V5MG7MFF.

Ganz, P., and C. Faunce. 2019. An evaluation of methods used to predict commercial fishing effort in

    Alaska. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-395, 19 p. doi.org/10.25923/gnyj-

    f281.

Guthrie III, C. M., Hv. T. Nguyen, M. Marsh, J. T. Watson, and J. R. Guyon. 2019. Genetic stock

    composition analysis of the Chinook salmon bycatch samples from the 2017 Bering Sea trawl

    fisheries. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-391, 36 p.

    doi.org/10.25923/dt1b-q428.

Guthrie III, C. M., Hv. T. Nguyen, M. Marsh, and J. R. Guyon. 2020. Genetic stock composition analysis

    of the Chinook salmon bycatch samples from the 2018 Gulf of Alaska trawl fisheries. U.S. Dep.

    Commer., NOAA Tech. Memo. NMFS-AFSC-405, 33 p. doi.org/10.25923/sw7t-gg49.

Hill, M.O. 1973. Diversity and evenness: A unifying notation and its consequences. Ecology 61: 225-236.

Nelson Jr., R., R. French, R. and J. Wall. 1981. Sampling by U.S. observers on foreign fishing vessels in the eastern Bering Sea and Aleutian Island region, 1977-78. Mar. Fish. Rev. 43:1-19.

NMFS (National Marine Fisheries Service). 2019a. *Draft* 2020 Annual Deployment Plan for Observers and Electronic Monitoring in the Groundfish and Halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/draft-2020-annual-deployment-plan-observers-and-electronic-monitoring-groundfish.

NMFS. 2019b. 2020 Annual Deployment Plan for Observers and Electronic Monitoring in the Groundfish and Halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/2020-annual-deployment-plan-observers-groundfish-and-halibut-fisheries-alaska.

NMFS. 2018. 2019 Annual Deployment Plan for Observers in the Groundfish and Halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/2019-annual-deployment-plan-observers-groundfish-and-halibut-fisheries-alaska.

NMFS. 2017. Final Environmental Assessment/ Regulatory Impact Review for Amendment 114 to the Fishery Management Plan for Groundfish of the Bering Sea and Aleutian Islands Management Area and Amendment 104 to the Fishery Management Plan for Groundfish of the Gulf of Alaska, and Regulatory Amendments: Analysis to Integrate Electronic Monitoring into the North Pacific Observer Program. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/ea-rir-amendment-114-fmp-groundfish-bsai-and-amendment-104-fmp-groundfish-goa-and

NMFS. 2013. 2014 Annual deployment plan for observers in the groundfish and halibut fisheries off

    Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska

    99802. Available at https://alaskafisheries.noaa.gov/sites/default/files/adp2014.pdf .

Wall J., French, R., and R. Nelson Jr. 1981. Foreign fisheries in the Gulf of Alaska, 1977-78. Mar. Fish.

    Rev. 43:20-35.

TABLES

Table 1. -- Comparison between predicted and actual trip days for partial coverage strata in 2019. Predicted values come from the 2019 Annual Deployment Plan (ADP).

| | Trip Days | | Difference | |
Strata	Predicted	Actual	Actual	Percent
HAL	8,561	9,426	865	10.1
POT - No Tender	2,468	2,421	-47	-1.9
POT - Tender	270	483	213	78.9
TRW - No Tender	4,759	4,167	-592	-12.4
TRW - Tender	151	332	181	119.9
Total	16,209	17,211	1,002	6.2

Table 2. -- Trip cancellation rates in the ODDS for 2019. A trip is cancelled by the system if the user did not identify whether fishing had occurred by the end of the year. "Paper" indicates that a trip was logged when the ODDS was not available.

Strata	Random number outcomes	Logged (*a*)	Cancelled by system (*b*)	Trips remaining (*c = a-b*)	Cancelled by user (*d*)	Paper	% User cancellation (*d/c* * 100)
HAL	Not Selected	1,552				0	
HAL	Selected	346	3	343	90	0	26.2
EM HAL	Not Selected	608				0	
EM HAL	Selected	312	0	312	14	0	4.5
POT - No Tender	Not Selected	499				0	
POT - No Tender	Selected	84	0	84	19	0	22.6
POT - Tender	Not Selected	118				0	
POT - Tender	Selected	14	0	14	3	0	21.4
EM POT	Not Selected	105				0	
EM POT	Selected	52	0	52	1	0	1.9
TRW - No Tender	Not Selected	1,321				0	
TRW - No Tender	Selected	426	0	426	88	0	20.7
TRW - Tender	Not Selected	46				0	
TRW - Tender	Selected	30	0	30	8	0	26.7
Total	Not Selected	4,249				0	
Total	Selected	1,264	3	1,261	223	0	17.7

Table 3. -- Number of remaining trips after cancellation in each trip-selection stratum that were selected using the initial random number generator (Random Number Selection) and those that remained after user manipulation (Total Final Selected). The relative impact of waivers in trip-selection is also shown (% Reduction of Selected Trips due to Waivers). **Not from random numbers.

Strata	Total Trips	Random number selection (*r*)	Inherited selection** (*i*)	Randomly selected but waived (*w*)	Total final selected (*T=r+i-w*)	% Selected from inherits ((i/T)*100)	% Reduction of selected trips due to waivers ($w/(T+w)$*100)
HAL	1,500	253	61	7	307	19.9	2.2
EM HAL	888	298	12	1	309	3.9	0.3
POT - No Tender	497	65	14	4	75	18.7	5.1
POT - Tender	103	11	4	0	15	26.7	0.0
EM POT	149	51	2	0	53	3.8	0.0
TRW - No Tender	1,528	338	50	0	388	12.9	0.0
TRW - Tender	58	22	6	1	27	22.2	3.6
Total	4,723	1,038	149	13	1,174	12.7	1.1

Table 4. -- Number of logged trips in each partial coverage stratum that were selected using the initial random number generator (Initial Random Selection) and those that remained after user manipulation (After Cancellations). The relative impact of inherits and waivers in trip-selection is also shown (With Inherits, After Waivers).

Strata	Trip disposition	Selected trips	Total trips	Actual selection (%)	Programmed selection (%)	p-value (H0: Actual = Programmed)
HAL	Initial Random Selection, a	346	1,898	18.23	17.71	0.548
	After Cancellations, b (a-b)	253	1,500	16.87	17.71	0.417
	With Inherits, c (a-b+c)	314	1,500	20.93	17.71	0.001
	After Waivers, d (a-b+c-d)	307	1,500	20.47	17.71	0.006
EM HAL	Initial Random Selection, a	312	920	33.91	30.00	0.011
	After Cancellations, b (a-b)	298	888	33.56	30.00	0.023
	With Inherits, c (a-b+c)	310	888	34.91	30.00	0.002
	After Waivers, d (a-b+c-d)	309	888	34.80	30.00	0.002
POT - No Tender	Initial Random Selection, a	84	583	14.41	15.43	0.528
	After Cancellations, b (a-b)	65	497	13.08	15.43	0.153
	With Inherits, c (a-b+c)	79	497	15.90	15.43	0.756
	After Waivers, d (a-b+c-d)	75	497	15.09	15.43	0.901
POT - Tender	Initial Random Selection, a	14	132	10.61	16.11	0.097
	After Cancellations, b (a-b)	11	103	10.68	16.11	0.178
	With Inherits, c (a-b+c)	15	103	14.56	16.11	0.789
	After Waivers, d (a-b+c-d)	15	103	14.56	16.11	0.789
EM POT	Initial Random Selection, a	52	157	33.12	30.00	0.385
	After Cancellations, b (a-b)	51	149	34.23	30.00	0.283
	With Inherits, c (a-b+c)	53	149	35.57	30.00	0.152
	After Waivers, d (a-b+c-d)	53	149	35.57	30.00	0.152
TRW - No Tender	Initial Random Selection, a	426	1,747	24.38	23.70	0.500
	After Cancellations, b (a-b)	338	1,528	22.12	23.70	0.149
	With Inherits, c (a-b+c)	388	1,528	25.39	23.70	0.125
	After Waivers, d (a-b+c-d)	388	1,528	25.39	23.70	0.125
TRW - Tender	Initial Random Selection, a	30	76	39.47	27.12	0.020
	After Cancellations, b (a-b)	22	58	37.93	27.12	0.076
	With Inherits, c (a-b+c)	28	58	48.28	27.12	0.001
	After Waivers, d (a-b+c-d)	27	58	46.55	27.12	0.002

Table 5. -- Number of total vessels (*V*), sampled vessels (*v*), total trips (*N*), sampled trips (*n*) for each stratum in 2019. The expected coverage (Expected?) and 95% confidence interval columns are expressed as percentages of the total number of trips taken within each stratum.

Coverage	Strata	*V*	*v*	*N*	*n*	Coverage Expected	Coverage Realized	95% Confidence Interval lower limit	95% Confidence Interval upper limit	Expected?
Full	*Full*	161	161	3,343	3,338	100.0	99.9			No
Partial	*HAL*	318	172	1,744	307	17.7	17.6	15.8	19.5	Yes
Partial	*EM HAL*	138	103	916	291	30.0	31.8	28.8	34.9	Yes
Partial	*POT - No Tender*	73	45	528	74	15.4	14.0	11.2	17.3	Yes
Partial	*POT - Tender*	30	12	44	13	16.1	29.5	16.8	45.2	No
Partial	*EM POT*	21	20	165	60	30.0	36.4	29.0	44.2	Yes
Partial	*TRW - No Tender*	78	70	1,568	395	23.7	25.2	23.1	27.4	Yes
Partial	*TRW - Tender*	26	12	56	20	27.1	35.7	23.4	49.6	Yes
Gear-based Total		584	397	5,016	1,159		23.1			
Partial	*Zero Coverage*	393	0	2,005	0	0.0	0.0			Yes
Partial	*Zero EM Research*	4	0	29	0	0.0	0.0			Yes
Total	Total	1085	510	10,393	4,497		43.3% Trips; 47.0% Vessels			

Table 6. -- The number of EM hard drives received and reviewed by gear type and month.

Strata	Data reviewed?	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Total
EM HAL	Yes	14	19	28	47	39	27	23	30	39	31	9	0	306
EM HAL	No	0	1	0	0	2	0	0	0	1	1	0	0	5
EM POT	Yes	18	0	0	5	1	0	2	5	10	5	3	1	50
EM POT	No	1	0	0	0	0	0	0	1	0	0	0	1	3

Table 7. -- The mean number of days taken for fixed gear EM data review by gear type. Columns are not additive, and instead represent two different ways of measuring review time, starting from either the end of the trip or from the date at which the hard drive was received.

Strata	Mean number of days between end of trip and data exported to NMFS	Mean number of days between hard drive received and data exported to NMFS
EM HAL	63	58
EM POT	92	79

Table 8. -- The number of *TRW - No Tender* pollock deliveries by port and coverage category.

FMP	Coverage category	Port	Total deliveries (*N*)	Observed deliveries (*n*)	% Observed
Bering Sea	Full	Akutan	831	831	100.0
Bering Sea	Full	Dutch Harbor	1,170	1,170	100.0
Bering Sea	Full	King Cove	90	90	100.0
Bering Sea	Full	Sand Point	1	1	100.0
Total	Full		2,092	2,092	100.0
Gulf of Alaska	Partial	Akutan	66	15	22.7
Gulf of Alaska	Partial	Dutch Harbor	1	1	100.0
Gulf of Alaska	Partial	King Cove	8	4	50.0
Gulf of Alaska	Partial	Kodiak	801	195	24.3
Gulf of Alaska	Partial	Sand Point	302	64	21.2
Total	Partial		1,178	279	23.7

Table 9. -- Number of trips by observation status in the 2019 trip-selection strata.

Strata	Observed	Unobserved
HAL	307	1,437
EM HAL	291	625
POT - No Tender	74	454
POT - Tender	13	31
EM POT	60	105
TRW - No Tender	395	1,173
TRW - Tender	20	36

Table 10. -- Results of permutation tests between monitored and unmonitored trips in the 2019 trip-selection strata. OD: Observed difference (monitored - unmonitored). A Bonferroni adjustment has been applied to p-values.

| Strata	Metric	NMFS areas	Days fished	Vessel length (ft)	Species landed	pMax species	Landed catch (t)
HAL	Observed difference	0.011	-0.662	0.849	-0.019	0.000	-0.905
	OD (%)	0.996	-12.334	1.530	-0.520	-0.056	-13.636
	p-value	1.000	< 0.001	1.000	1.000	1.000	0.030
EM HAL	Observed difference	-0.003	-0.309	-1.344	0.525	-0.031	0.206
	OD (%)	-0.294	-6.224	-2.589	13.380	-3.514	2.951
	p-value	1.000	0.252	0.204	< 0.001	0.054	1.000
POT - No Tender	Observed difference	-0.011	-0.429	0.812	-0.055	0.012	4.153
	OD (%)	-1.048	-9.360	1.116	-2.849	1.188	13.241
	p-value	1.000	0.870	1.000	1.000	0.156	1.000
POT - Tender	Observed difference	-0.060	4.181	8.417	0.218	-0.001	175.762
	OD (%)	-4.679	38.089	9.203	7.279	-0.074	100.077
	p-value	1.000	0.144	1.000	1.000	1.000	< 0.001
EM POT	Observed difference	-0.019	-0.719	0.357	0.486	-0.392	-1.732
	OD (%)	-1.882	-16.757	0.490	21.202	-31.701	-6.952
	p-value	1.000	0.144	1.000	0.012	1.000	1.000
TRW - No Tender	Observed difference	-0.046	-0.033	0.859	-0.733	0.011	-3.312
	OD (%)	-4.352	-1.238	1.013	-11.856	1.235	-3.501
	p-value	< 0.001	1.000	1.000	0.024	1.000	1.000
TRW - Tender	Observed difference	0.172	0.578	5.528	-0.028	0.013	86.389
	OD (%)	15.811	9.746	8.135	-0.615	1.281	66.601
	p-value	0.312	1.000	1.000	1.000	1.000	0.330
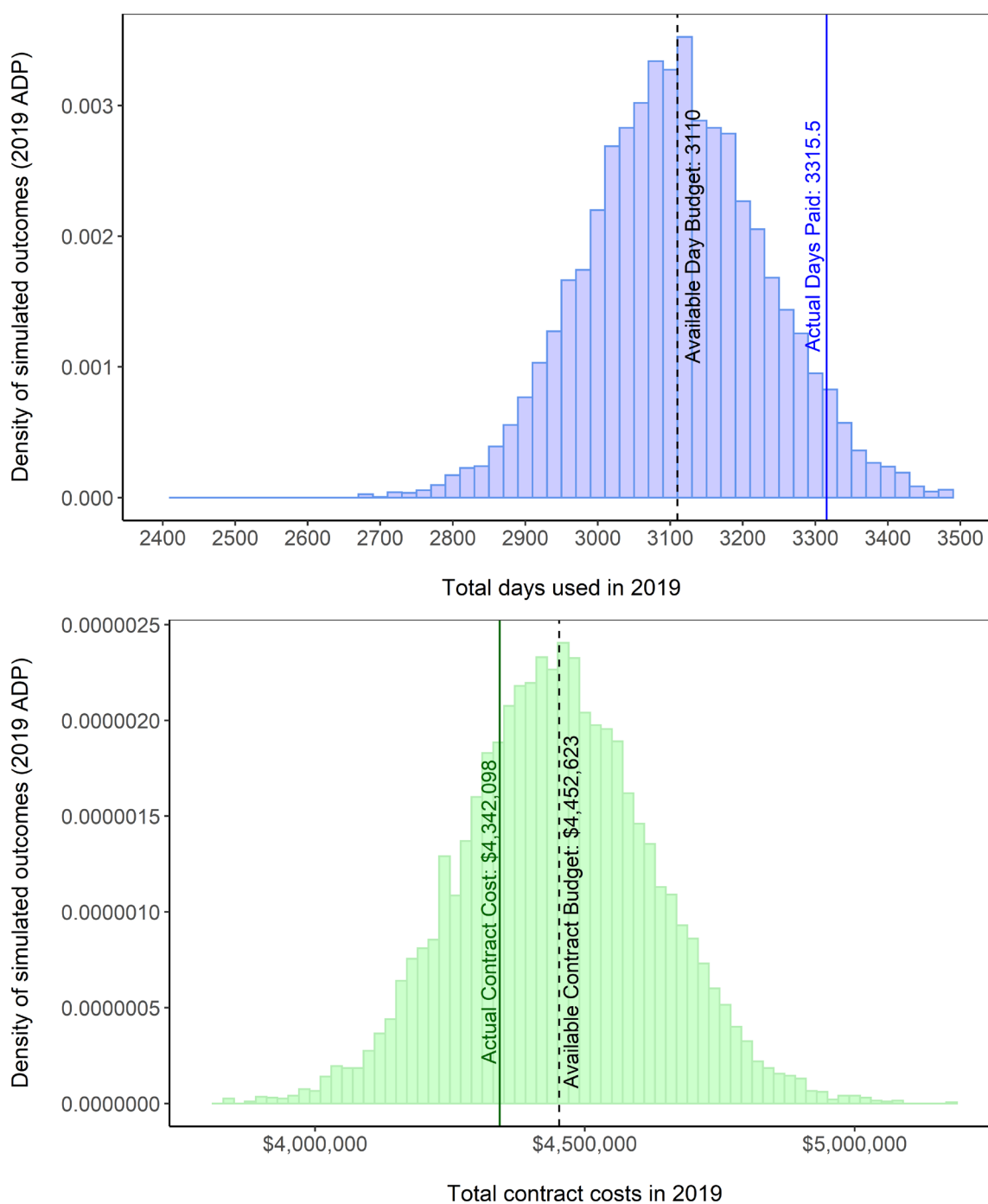
Figure 1. -- Total number of observer sea days purchased (top panel) and total cost of observing those sea days (bottom panel). Vertical bars signify the range of potential outcomes predicted by the 2019 Annual Deployment Plan. Dashed lines signify expected outcomes. Solid lines signify what actually occurred in 2019.
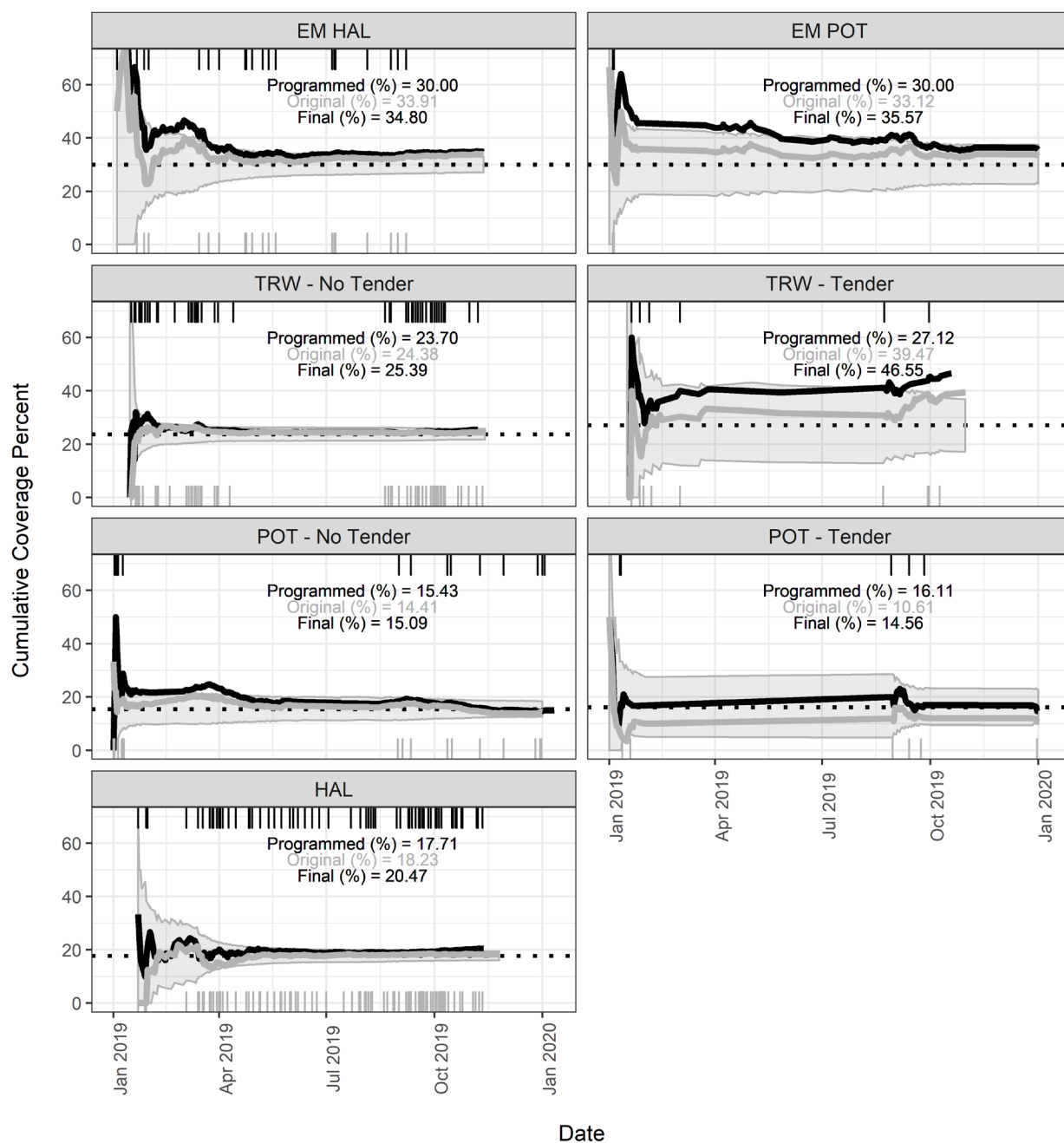
Figure 2. -- Rate of selected trips logged into ODDS organized by original date entered for all trips (grey line and grey text), and final date considering only non-cancelled trips (black line and black text). The programmed selection rate is depicted as the dotted line. Grey shaded areas denote the range of coverage rate corresponding to the 95% confidence intervals expected from the binomial distribution. The final coverage rates were higher than if trip dates had not been altered and/or cancelled. Vertical tick marks on the horizontal axis depict dates when an ODDS trip was selected due to a prior trip being cancelled that was selected for observer coverage (grey on the bottom for originally logged trips, and black on the top for trips after user manipulation).
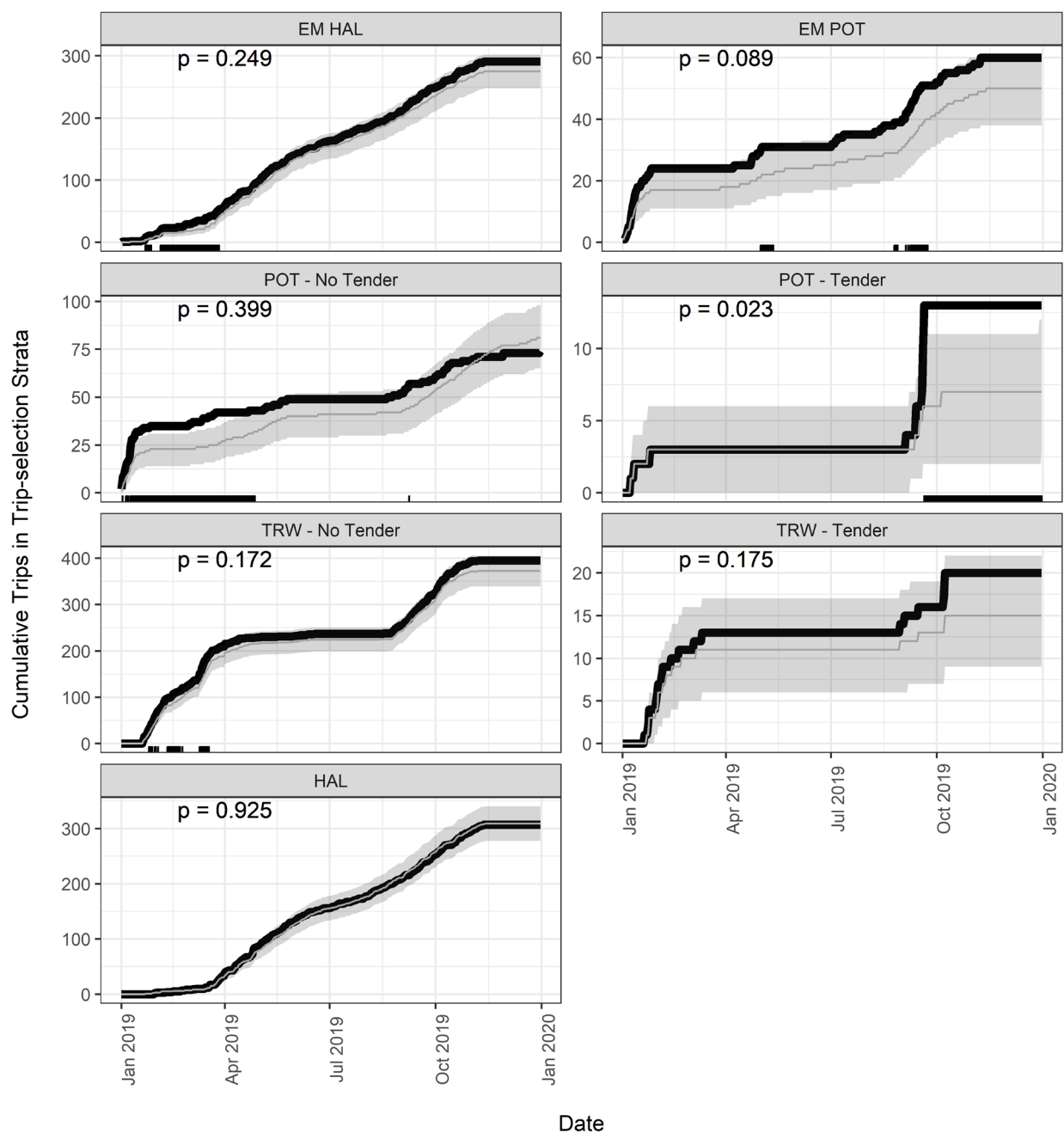
Figure 3. -- Cumulative number of trips monitored during 2019 (black line) compared to the expected range of observed trips (shaded area) given fishing effort and sampling rates. Dates where the monitored number of trips is outside of expected (less or more than the range) are depicted as tick marks on the horizontal x-axis. The results of tests that the observed rate at the end of the year derived from a binomial distribution sampled at the selection rate are denoted as *p*-values.
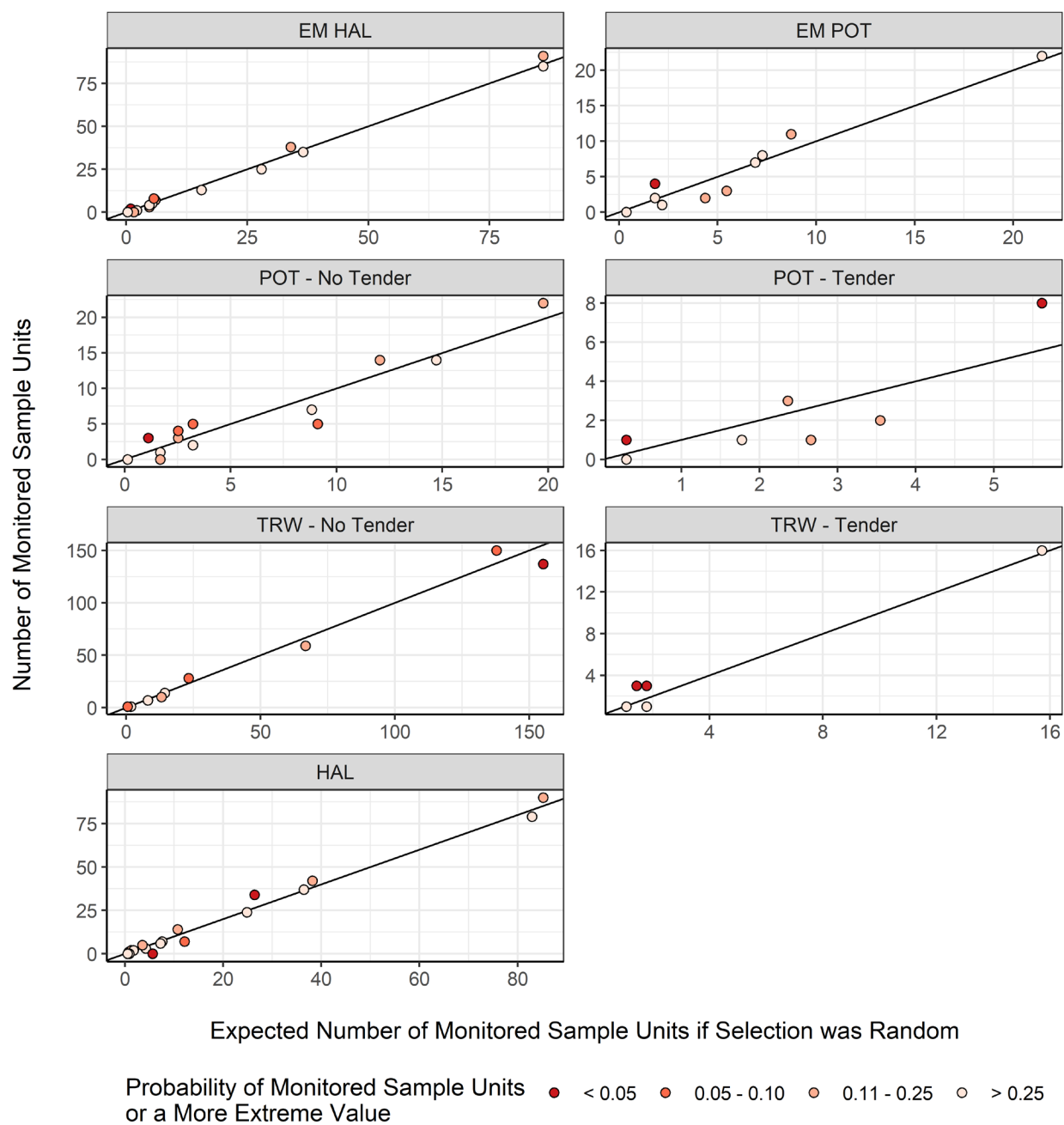
Figure 4. -- Comparison plots depicting the number of monitored sample units compared to the number of expected monitored sample units for each partial coverage stratum. Each point on a plot represents a NMFS Area. The darker the point, the more unusual the result.
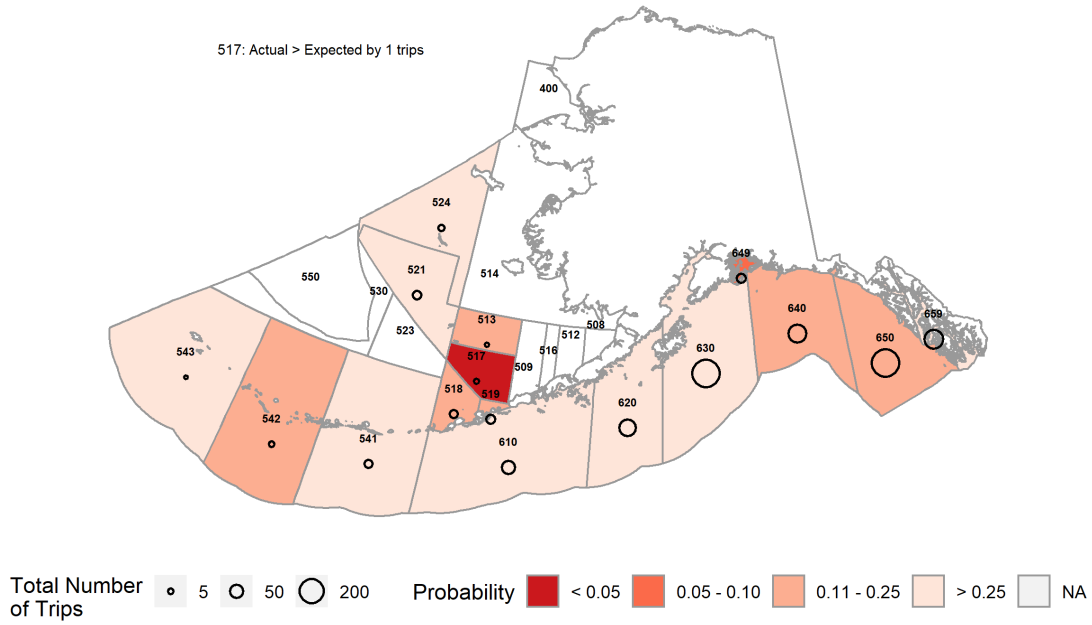
EM HAL 2019

517: Actual > Expected by 1 trips

Total Number of Trips: ● 5 ◯ 50 ◯ 200   Probability: ▪ < 0.05 ▪ 0.05 - 0.10 ▪ 0.11 - 0.25 ▪ > 0.25 ▫ NA

Figure 5. -- Probability of monitoring the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *EM HAL* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

EM POT 2019

512: Actual > Expected by 2 trips

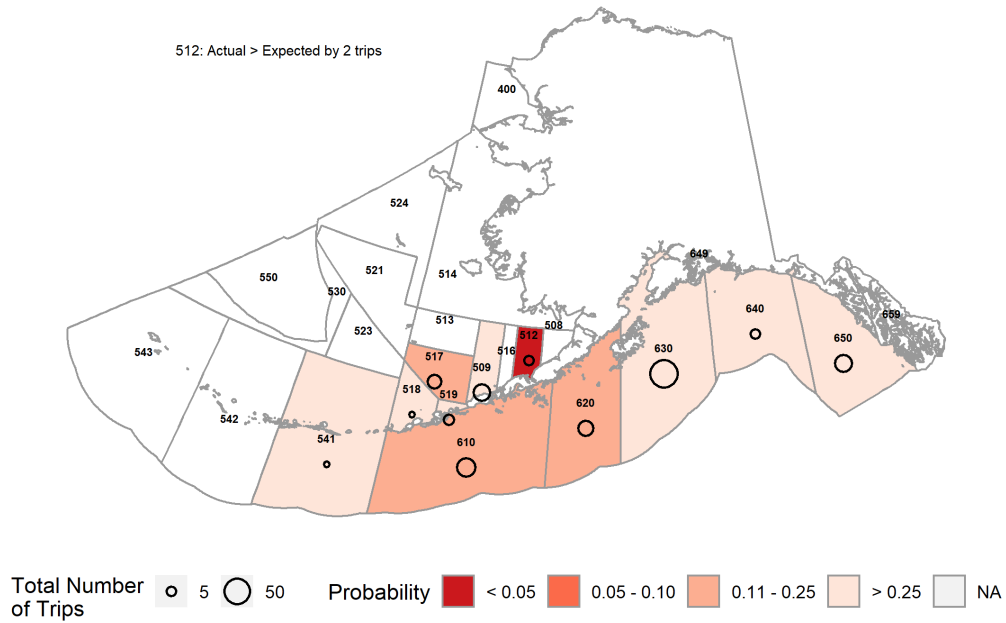Total Number of Trips: ● 5 ◯ 50   Probability: ▪ < 0.05 ▪ 0.05 - 0.10 ▪ 0.11 - 0.25 ▪ > 0.25 ▫ NA

Figure 6. -- Probability of monitoring the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *EM POT* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

HAL 2019

542: Actual < Expected by 6 trips
610: Actual > Expected by 8 trips

Total Number of Trips: • 5 ○ 50 ◯ 200

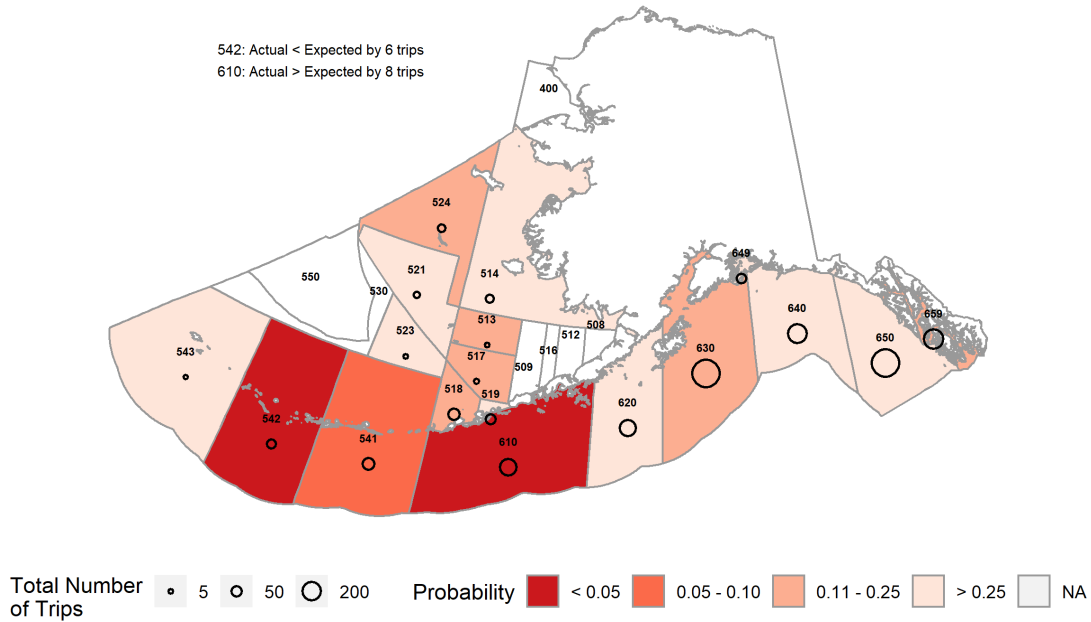Probability: ■ < 0.05 ■ 0.05 - 0.10 ■ 0.11 - 0.25 ■ > 0.25 □ NA

Figure 7. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *HAL* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

POT - No Tender 2019

518: Actual > Expected by 2 trips

Total Number of Trips: • 5 ◯ 50

Probability: ■ < 0.05 ■ 0.05 - 0.10 ■ 0.11 - 0.25 ■ > 0.25 □ NA

Figure 8. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *POT - No Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.
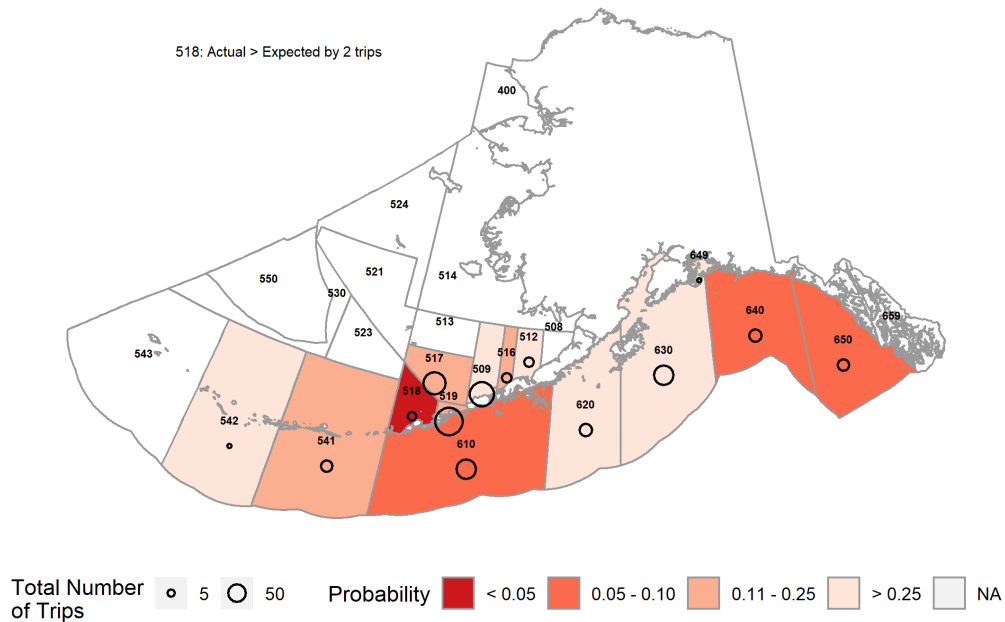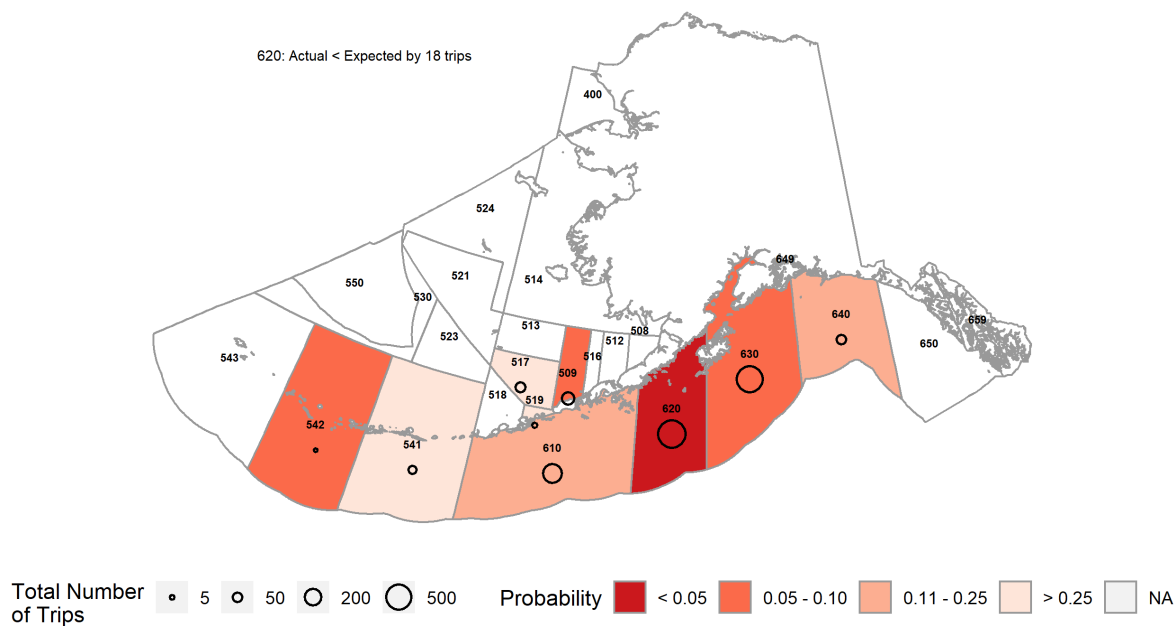
TRW - No Tender 2019

Figure 9. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *TRW - No Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

POT - Tender 2019

Figure 10. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *POT - Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.
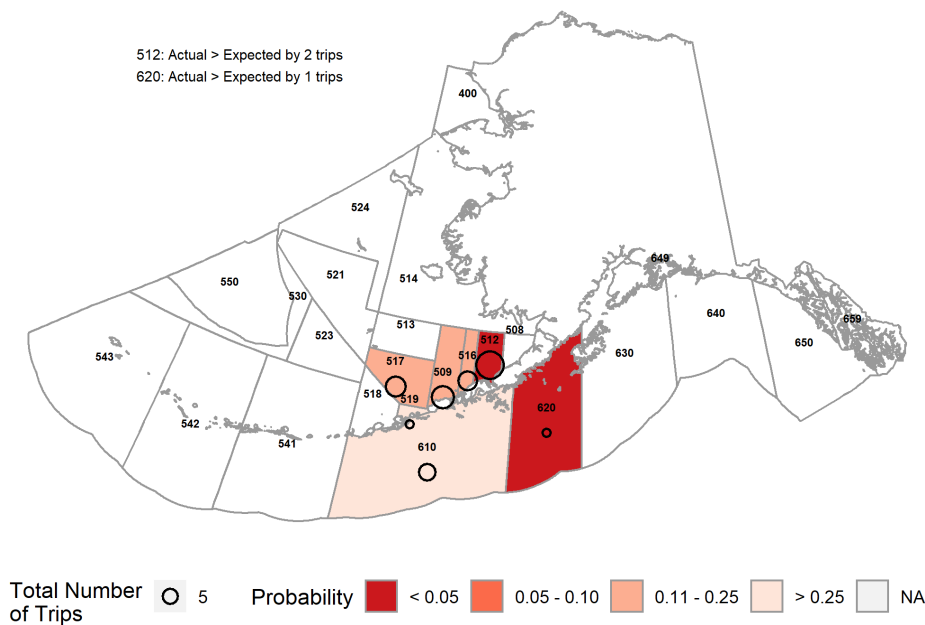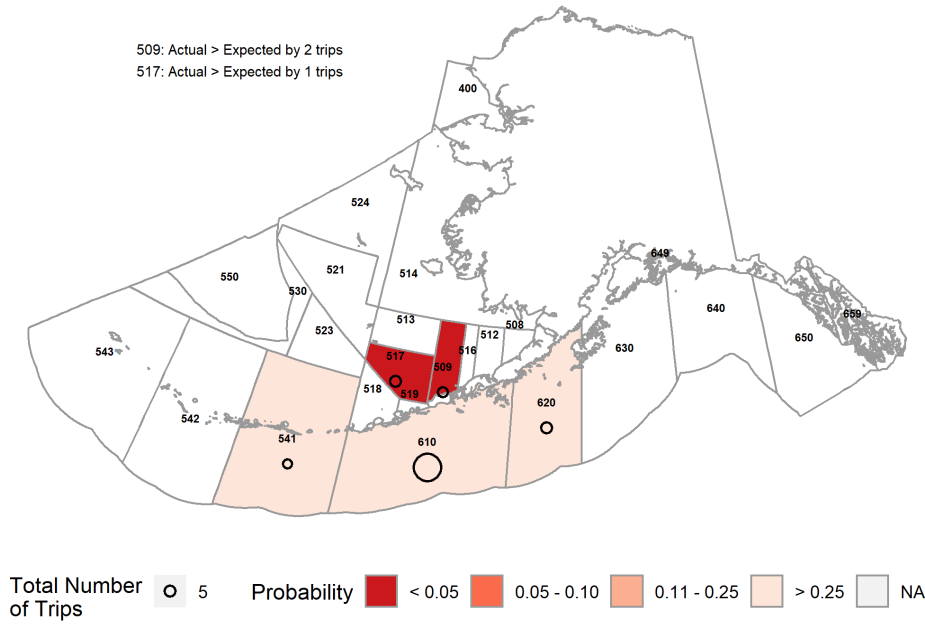
Figure 11. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *TRW - Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.
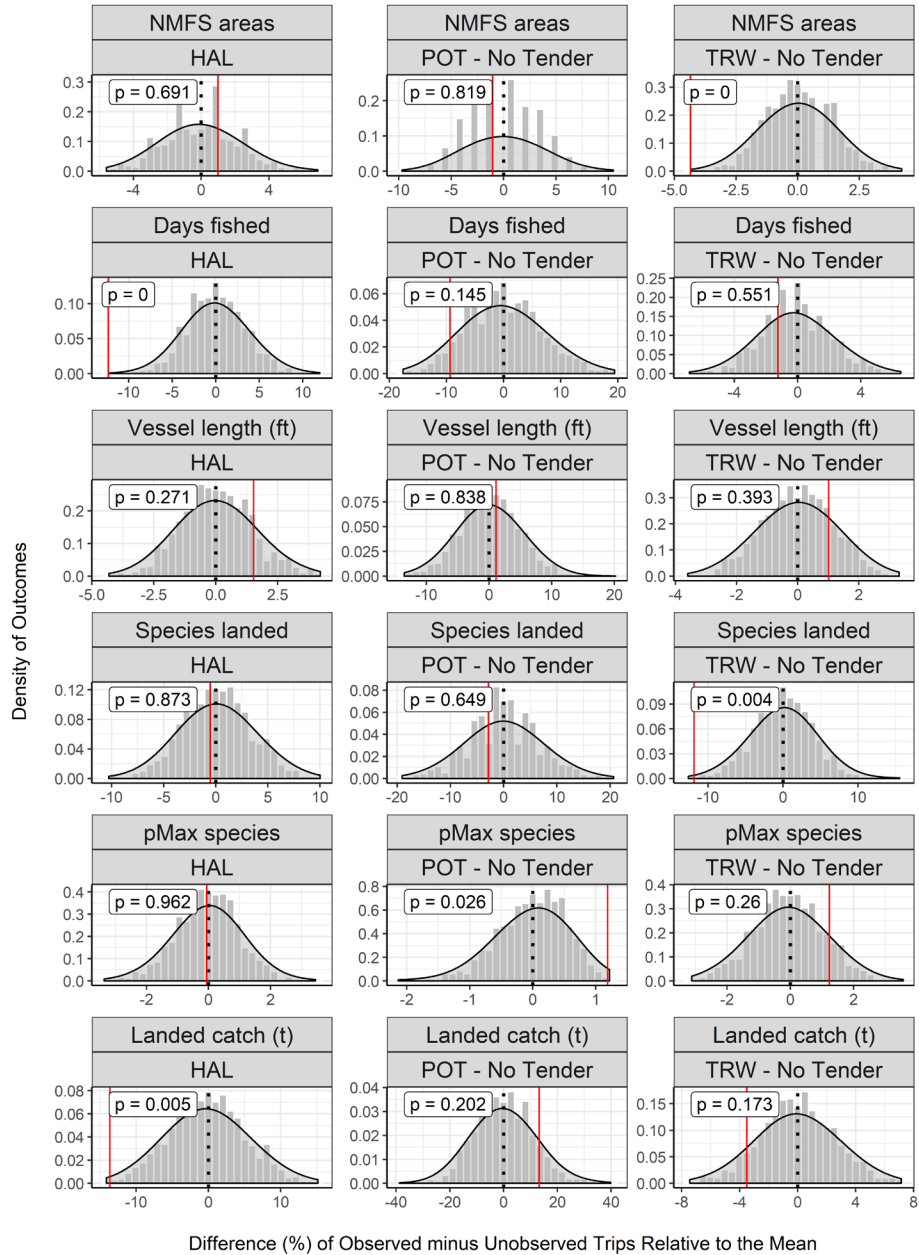
Figure 12. -- Example of results from permutation tests depicting percent differences between observed and unobserved trips for observer pool strata in the partial coverage category. Grey bars depict the distribution of differences between observed and unobserved trips where the assignment of observed status has been randomized (this represents the sampling distribution under the null hypothesis that observed and unobserved trips are the same). The vertical red line denotes the actual difference between observed and unobserved trips. Values on the x-axis have been scaled to reflect the relative (%) differences in each metric. The *p*-value for each test is denoted in the upper left corner. Low *p*-values are reason to reject the null hypothesis and conclude that there is an observer effect. Results from all permutation tests can be found in the Tables section of this report.
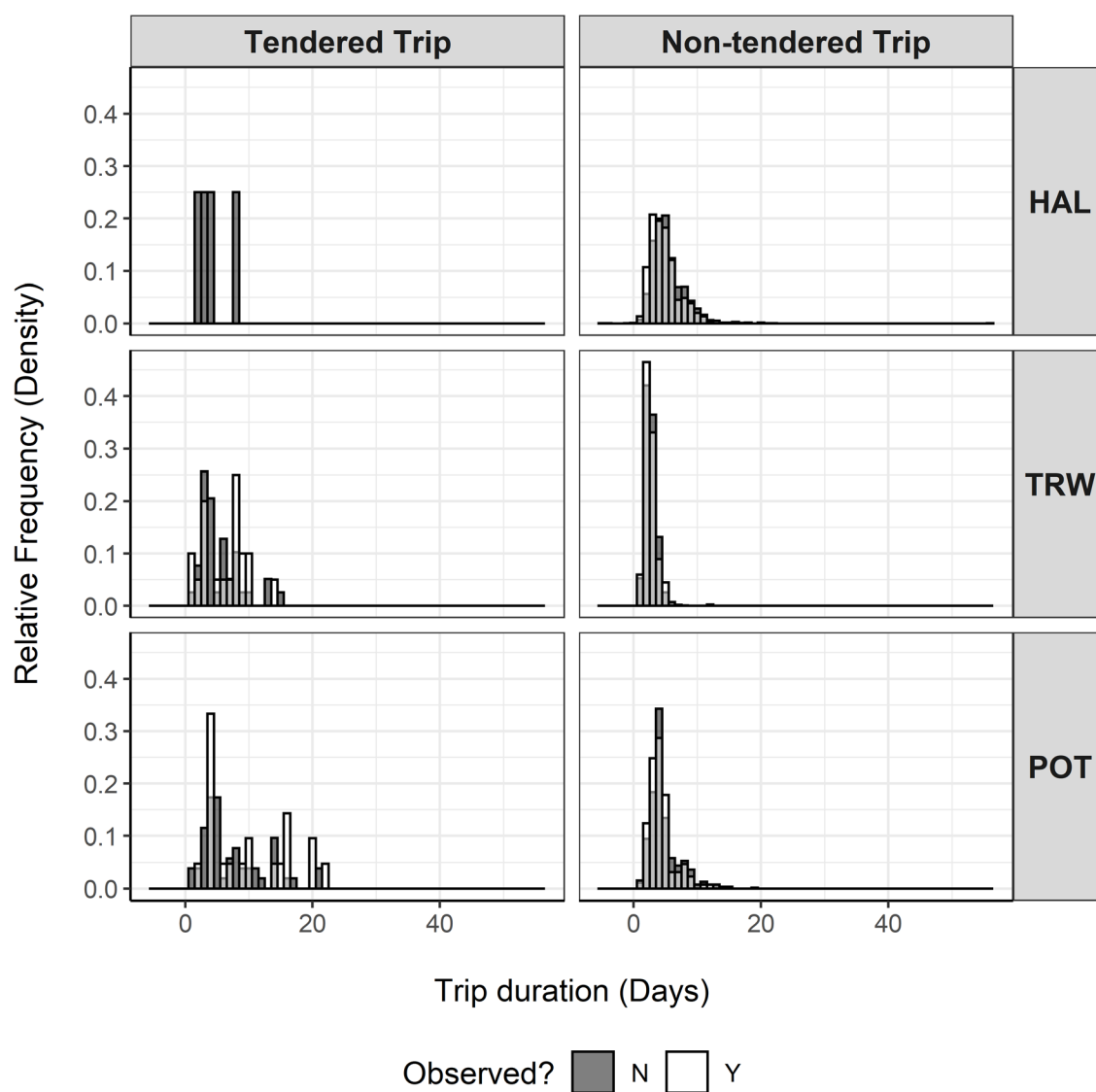
Figure 13. -- Distribution of trip durations for vessels in the partial coverage category by gear and observation status. Observed trips are depicted as transparent white bars overtop of solid black bars for unobserved trips. Trip durations where both observed and unobserved status exist are depicted in gray (This is not the same as a stacked bar chart, in which the height of the bar would reflect observed and unobserved on top of one another- this plot has each observation status in front of the other).
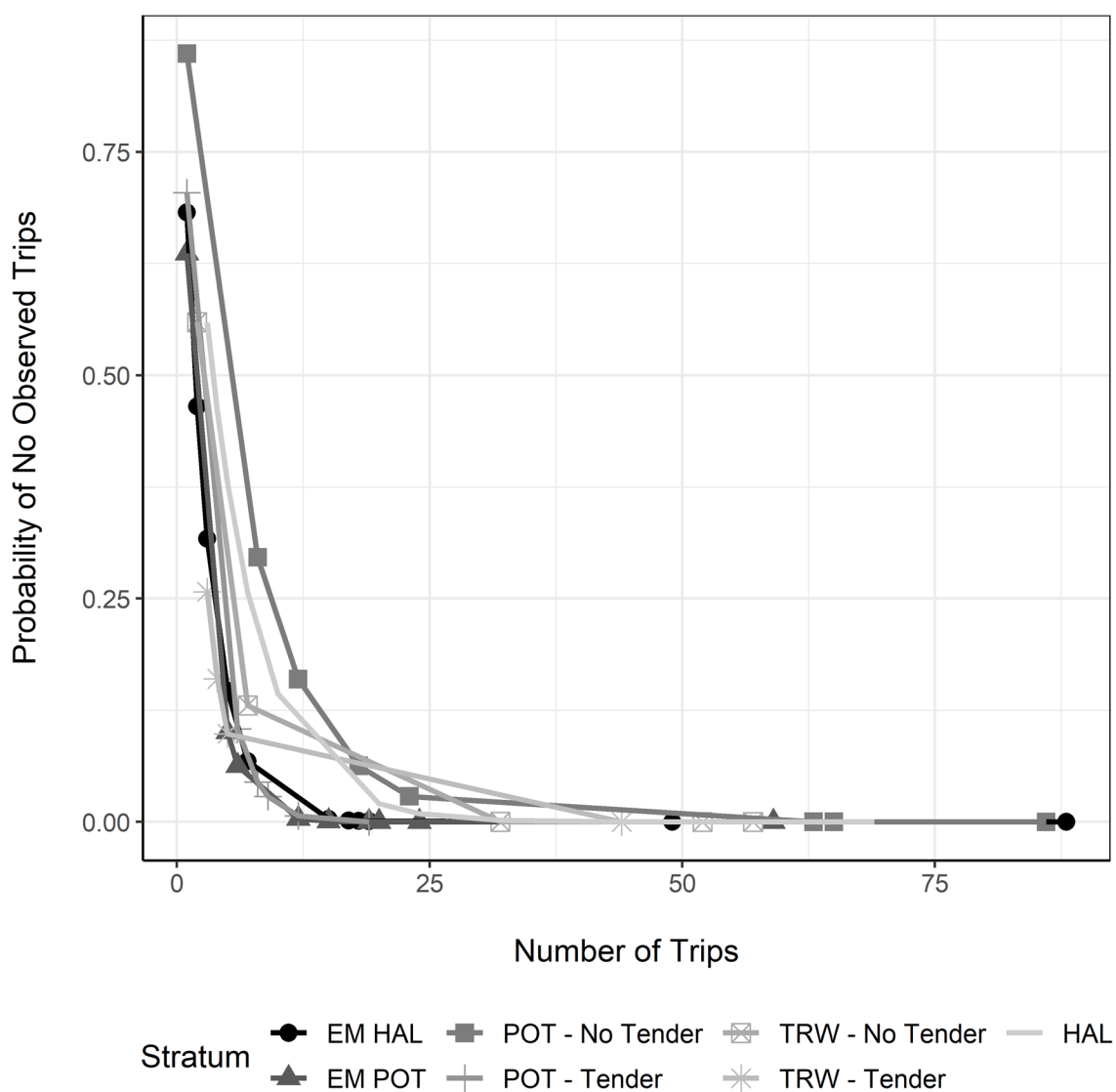
Figure 14. -- Probability of monitoring no trips in a NMFS Area and stratum given fishing effort and sampling rate. The x-axis has been truncated to increase resolution at low levels of fishing effort. The likelihood of having no monitoring data decreases with increasing total fishing effort and selection rate. The selection rate is 17.71% in the *HAL* stratum, 15.43% in the *POT - No Tender* stratum, 16.11% in the *POT - Tender* stratum, 23.70% in the *TRW - No Tender* stratum, 27.12% in the *TRW - Tender* stratum, 30.00% in the *EM HAL* stratum, and 30.00% in the *EM POT* stratum.

APPENDIX A – EVALUATION OF PELAGIC AND NON-PELAGIC TRAWL TRIPS

INTRODUCTION

At its June 2017 meeting, the North Pacific Fishery Management Council (Council) requested that the National Marine Fisheries Service (NMFS) evaluate whether there is evidence of an observer effect in either pelagic trawl (PTR) or non-pelagic trawl (NPT) gear fished by partial coverage vessels (AFSC and AKRO 2018, p. 54). These two gear types are typically used for different styles of fishing, with NPT gear associated with bottom contact and PTR gear typically fished in the water column. The Council's request followed a Fishery Monitoring Advisory Committee (FMAC) request for the evaluation, including a discussion about the "pros and cons" of separate observer deployment strata for those two gear types. The NMFS performed the requested analyses, and the resulting recommendation was to not separate the trawl gears into two separate strata (AFSC and AKRO 2018, Appendix A; AFSC and AKRO 2019, Appendix A). Following these initial analyses, the FMAC expressed interest in continuing to see an evaluation of the NPT and PTR gear types. The analysis presented here is intended to serve as that continued evaluation.

Although the North Pacific Observer Program (Observer Program) does not currently deploy observers into separate NPT and PTR strata, the Catch Accounting System (CAS) post-stratifies observer and landings data based on whether the trip is recorded as NPT or PTR on the landing report ("fish ticket") or in the observer data. The fact that trawl trips are post-stratified by NPT and PTR gear means that estimates of bycatch for unobserved NPT trips are based solely on observed NPT trips (not PTR trips), and estimates of bycatch for unobserved PTR trips are based solely on observed PTR trips (not NPT trips). In both cases, the vessel operator reports the gear type being used to the processor. On observed trips, observers are expected to verify the reported gear type. Regulations at 50 CFR 679.2 (definitions) define NPT and PTR gear to be of certain configurations (e.g., floats, mesh configurations, line configurations).

The NPT and PTR gear types are associated with different fishery management issues, with salmon bycatch being the primary issue for the PTR pollock fisheries, and halibut PSC being of concern for some NPT fisheries. Being a relatively rare bycatch species in the PTR pollock fisheries, salmon are accounted for shoreside when an observer is on board a vessel that is targeting pollock and not delivering to a tender. In contrast, halibut discard estimates are based on samples collected by observers at sea. In both cases, data from observed trips are used to make estimates for unobserved trips, but at-sea observer samples are inherently more variable than the shoreside census conducted for salmon in pollock fisheries. Because of this sampling dynamic, and the differing incentives for different fisheries, a concern raised by some stakeholders has been that vessels selected for observer coverage are disproportionately opting to fish for pollock instead of species that are typically fished with NPT gear. Such behavior would result in higher observer coverage in PTR gear since it is used to target pollock.

In contrast to analyses presented elsewhere in this document, multiple years of data are used in this appendix, since significance tests on observation rates have never before been conducted within the NPT and PTR gear types. Significance tests rely on the hypergeometric distribution which, when estimating the probability of observing a given number of NPT or PTR trips, accounts for the total number of observed non-tendered trawl trips that occurred. Therefore, a significant result within a gear type means that the number of observed trips was significantly different than the number of observed trips that were expected within that gear type, given the total number of observed non-tendered trawl trips that occurred. As in the main section of this report, a Bonferroni adjustment has been applied to all permutation test p-values in order to control for multiple comparisons. This adjustment was not applied to permutation test p-values presented in last year's annual report (AFSC and AKRO 2019). This adjustment corrects for the increased probability of detecting a false positive result due to conducting multiple tests on the same data. One drawback of this adjustment is the decreased ability to detect true differences if they exist.

Separate from differing coverage levels between gear types, the original request made by the Council was to evaluate whether or not there is evidence of an observer effect within PTR and NPT

fisheries (AFSC and AKRO 2018, p. 54). We first responded to that request by providing the results of permutation tests that measured differences between observed and unobserved trips (AFSC and AKRO 2019, Appendix A). Evidence of observer effects within non-tendered trawl trips has been shown in multiple Annual Reports (AFSC and AKRO 2017, AFSC and AKRO 2018, AFSC and AKRO 2019), so one motivation for performing permutation tests within gear type is to give more granularity to those stratum-level results. All analyses in this appendix consider only non-tendered trips.

RESULTS

Since 2016, 99.7% of the partial coverage category PTR landings targeted pollock (Appendix Table A-1). Of these 5,425 pollock trips, 96.6% had a catch composition of at least 95% pollock, which falls into the CAS "pelagic" pollock target (suggesting midwater tows). The remaining pollock landings were in the "bottom" pollock target category, which is based on the pollock being the predominant species retained (but less than 95% of the retained catch). The predominant targets for vessels fishing NPT gear were Pacific cod (49.7% of trips) and arrowtooth flounder (34.4% of trips), followed by pollock (9.4% of trips; Appendix Table A-1).

Observation rates for PTR gear were significantly higher than expected in one of the four years analyzed here (Appendix Table A-2). Observation rates for NPT gear were significantly lower than expected in two of the four years analyzed (Appendix Table A-2). Mixed-gear trips, during which the vessel fishes both NPT and PTR gear, were not uncommon (Appendix Table A-2).

The majority of permutation tests conducted show no significant difference between observed and unobserved trips (Appendix Table A-3). Of the significant differences that did occur, most occurred in only one year for any given metric and gear type combination (Appendix Table A-3). Two differences were significant in more than one year: observed NPT trips landed fewer species (three of the four years tested) and less catch (two of the four years tested) than unobserved NPT trips (Appendix Table A-3). In 2019, two metrics showed significant differences: NPT trips landed an average of two fewer species when

61

observed, and PTR fished an average of 0.03 fewer NMFS areas when observed (Appendix Table A-3). There were no significant differences in 2019 between observed and unobserved trips in the number of days fished, vessel length, proportion of catch that is made up of the predominant species (pMax), or amount of landed catch (Appendix Table A-3).

DISCUSSION

While it was known prior to these analyses (AFSC and AKRO 2018, Appendix A; AFSC and AKRO 2019, Appendix A) that NPT and PTR target different species, it was not known just how much observation rates and observer effects differ between the two gear types. Results presented here suggest that observation rates differed from what was expected in some years, but not others (Appendix Table A-2). Although the significant differences occurred in the two years with the lowest observation rates of non-tendered trawl trips, we do not test for a significant relationship between non-tendered trawl observation rates and differences between observation rates within the two gear types.

Considering all years for which permutation tests have been performed within the NPT and PTR gear types, there is no clear pattern over time in terms of which metrics show an observer effect (Appendix Table A-3). Regardless, it's important to note that creating separate NPT and PTR strata would not change the feature of fisheries monitoring that allows for observer effects in the first place: the ability of vessels to behave differently on observed trips compared to unobserved trips. Stratification does, however, have the potential to influence whether or not gear types are observed at expected rates. Although this analysis included a significance test for observation rates within the NPT and PTR gear types, the most recent Annual Deployment Plan has one trawl stratum that includes both NPT and PTR trips (NMFS 2019b). This means that, while we can analyze whether NPT and PTR were observed at expected rates given the number of observed trawl trips that occurred and the number of trips that occurred within the NPT and PTR gear types, there is currently no enforceable expectation that vessels use a particular gear type on an observed trawl trip. Despite the potential for separate strata to give NMFS

more influence over whether or not the NPT and PTR gear types are observed at expected rates, we do not currently see evidence that such additional influence is warranted. In 2019, NPT, PTR, and mixed-gear trips were all observed at expected rates (Appendix Table A-2).

In addition to a lack of evidence to support stratification, there are logistical challenges to deploying observers into separate NPT and PTR strata. These challenges were identified in previous analyses (AFSC and AKRO 2018, Appendix A; AFSC and AKRO 2019, Appendix A), and they include the potential incentive for vessel operators to log trips under one gear type to obtain the more desirable selection rate, but then fish using the other gear type. A similar pattern has been seen with tender strata, in which vessels would log tender trips and then deliver shoreside, or vice-versa (AFSC and AKRO 2019). The inaccurate reporting of tender status was one reason the NMFS decided not to stratify by tender status in 2020 (NMFS 2019a, Appendix B). For all the above reasons, the NMFS has not created separate strata for the NPT and PTR gear types.

CITATIONS

AFSC (Alaska Fisheries Science Center) and AKRO (Alaska Regional Office). 2019. North Pacific Observer Program 2018 Annual Report. AFSC Processed Rep. 2019-04, 148 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. Available at http://www.afsc.noaa.gov/Publications/ProcRpt/PR2019-04.pdf

AFSC and AKRO. 2018. North Pacific Observer Program 2017 Annual Report. AFSC Processed Rep. 2018-02, 136 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. Available at http://www.afsc.noaa.gov/Publications/ProcRpt/PR2018-02.pdf

AFSC and AKRO. 2017. North Pacific Observer Program 2016 Annual Report. AFSC Processed Rep. 2017-07, 143 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. Available at http://www.afsc.noaa.gov/Publications/ProcRpt/PR2017-07.pdf

NMFS (National Marine Fisheries Service). 2019a. *Draft* 2020 Annual Deployment Plan for Observers and Electronic Monitoring in the Groundfish and Halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/draft-2020-annual-deployment-plan-observers-and-electronic-monitoring-groundfish.

NMFS. 2019b. 2020 Annual Deployment Plan for Observers and Electronic Monitoring in the Groundfish and Halibut Fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at https://www.fisheries.noaa.gov/resource/document/2020-annual-deployment-plan-observers-groundfish-and-halibut-fisheries-alaska.

Appendix Table A-1. -- Number of trips (N) by target species (Target) for NPT and PTR gear types (Gear) between 2016 and 2019. For the purpose of this table, mixed-gear trips are excluded.

Gear	Target	N
NPT	Pacific cod	1,178
	Arrowtooth flounder	816
	Pollock	223
	Flatfish (shallow water)	104
	Flathead sole	22
	Rex sole	7
	Sablefish	5
	Atka mackerel	4
	Rockfish	4
	Yellowfin sole	4
	Other	3
NPT Total		**2,370**
PTR	Pollock	5,425
	Arrowtooth flounder	6
	Flatfish (shallow water)	3
	Pacific cod	2
	Rockfish	2
	Atka mackerel	1
PTR Total		**5,439**

Appendix Table A-2. -- Number of total trips (N), sampled trips (n), and % observed for NPT and PTR gear type. Significance tests rely on the hypergeometric distribution, which accounts for the number of observed non-tendered trips that occurred when estimating the probability of observing a given number of NPT or PTR trips. For the purpose of this table, mixed-gear trips are counted separately from single-gear trips.

Gear	N	n	% Observed all trawl	% Observed by gear	p-value	Realized meets expected?
2016						
PTR	1,560	421	26.2	27.0	0.10	Yes
NPT	844	205	26.2	24.3	0.07	Yes
NPT & PTR	62	19	26.2	30.6	0.17	Yes
2017						
PTR	1,544	350	20.7	22.7	0.00	No
NPT	508	82	20.7	16.1	0.00	No
NPT & PTR	38	1	20.7	2.6	0.00	No
2018						
PTR	1,292	272	20.3	21.1	0.09	Yes
NPT	528	92	20.3	17.4	0.03	No
NPT & PTR	44	14	20.3	31.8	0.02	No
2019						
PTR	1,043	267	25.2	25.6	0.28	Yes
NPT	490	121	25.2	24.7	0.41	Yes
NPT & PTR	35	7	25.2	20.0	0.31	Yes

Appendix Table A-3. -- Results of permutation tests between observed and unobserved trips within the NPT and PTR gear types. For the purpose of these tests, mixed-gear trips are excluded. A Bonferroni adjustment has been applied to *p*-values.

Gear	Metric	NMFS areas	Days fished	Vessel length (ft)	Species landed	pMax species	Landed catch (t)
2016							
NPT	Observed difference	-0.037	-0.444	-1.044	-1.764	0.049	-23.684
NPT	OD (%)	-2.925	-11.499	-1.188	-26.854	6.022	-39.879
NPT	p-value	1.000	< 0.001	1.000	< 0.001	0.018	< 0.001
PTR	Observed difference	-0.012	0.193	4.884	-0.185	-0.001	7.952
PTR	OD (%)	-1.181	8.383	6.024	-3.833	-0.112	8.484
PTR	p-value	1.000	0.006	< 0.001	0.678	1.000	< 0.001
2017							
NPT	Observed difference	0.063	-0.143	1.504	-1.521	0.056	-16.168
NPT	OD (%)	5.051	-3.688	1.689	-21.083	6.765	-19.774
NPT	p-value	1.000	1.000	1.000	0.048	0.066	0.066
PTR	Observed difference	-0.012	-0.032	-1.437	-0.224	-0.002	-3.072
PTR	OD (%)	-1.178	-1.381	-1.698	-5.200	-0.169	-2.857
PTR	p-value	1.000	1.000	0.948	0.636	0.036	0.960
2018							
NPT	Observed difference	-0.089	-0.388	-4.309	-0.360	0.032	-18.648
NPT	OD (%)	-7.746	-10.588	-5.059	-4.035	4.142	-26.359
NPT	p-value	0.210	0.084	0.084	1.000	0.954	< 0.001
PTR	Observed difference	-0.001	0.064	-0.644	0.195	-0.002	-2.076
PTR	OD (%)	-0.144	2.496	-0.765	4.410	-0.158	-1.986
PTR	p-value	1.000	1.000	1.000	0.930	1.000	1.000
2019							
NPT	Observed difference	-0.069	-0.083	2.128	-2.024	0.023	-8.314
NPT	OD (%)	-6.091	-2.539	2.388	-21.215	3.111	-10.357
NPT	p-value	0.372	1.000	1.000	0.006	1.000	0.498
PTR	Observed difference	-0.034	0.012	0.427	0.022	-0.001	-0.668
PTR	OD (%)	-3.269	0.494	0.517	0.491	-0.055	-0.662
PTR	p-value	0.024	1.000	1.000	1.000	1.000	1.000

APPENDIX B – GAP ANALYSIS OF THE 2019 NORTH PACIFIC OBSERVER

PROGRAM

INTRODUCTION

This analysis evaluates the deployment of observers and electronic monitoring (EM) systems within the partial coverage category in the context of catch estimation. Catch estimation relies on representative sampling of fishing activity which is achieved through random deployment of monitoring coverage. Within the observer and EM pools, fishing trips are randomly selected for monitoring via the Observer Declare and Deploy System (ODDS) at strata-specific trip selection rates. In theory, random trip selection should result in proportionate deployment of sampling effort to all post-strata within each stratum, i.e., monitored trips are distributed similarly to all fishing effort spatially, temporally, and between fisheries. In reality, there are various factors that may cause sampling effort to be disproportional to fishing effort within a stratum and therefore may result in a lack of and/or non-representative samples from which to generate catch and discard estimates. Although observers and EM systems are not deployed into individual fisheries within a given stratum, by evaluating coverage within post-strata we can better understand some of the departures from expected deployment patterns found in the broader assessment presented in the main section of this report.

Although trip selection is a random process, the resulting sampling effort may not be proportionally distributed among post-strata due to random chance, cancellation policies in the ODDS, and/or observer effects. For example, monitoring coverage can be delayed by logging multiple trips, cancelling trips, and inheriting monitoring coverage from cancellation of selected trips. In addition, fishing activity may be influenced by selection status with monitored trips having different duration, location, or target species than unmonitored trips. These factors can potentially result in spatiotemporal differences between monitored trips and cause catch estimates to extrapolate from data pooled at coarser spatiotemporal scales.

In this appendix, we examine the patterns of observer and EM coverage relative to total fishing activity at a finer scale than presented in the main section of this report. Results here are intended to provide additional detail to some of the earlier findings, however, because these are post-hoc analyses being conducted at a finer scale than overall deployment specified in the 2019 ADP. Care should be taken when interpreting the results.

METHODS

The methods used in this analysis are similar to those employed in the gap analysis in Appendix C of the 2020 Draft Annual Deployment Plan (ADP, NMFS 2019). Partial coverage fishing effort data from 2019 was used in conjunction with a simplified version of the Catch Account System's (CAS) post-stratification process to quantify the degree to which data from monitored trips are available within specified spatiotemporal distances to unmonitored fishing trips. In general, the larger the distance, the greater the potential for problematic gaps (sparse or no data collected) within a given spatiotemporal bin (e.g., post-strata in CAS or data groupings used within stock assessments).

This analysis included four distinct types of monitoring coverage that are used within and between partial coverage selection pools: 1) Monitored observer pool trips relative to unmonitored observer pool trips (OB-OB), 2) Monitored observer pool trips relative to all zero-selection pool trips (OB-ZE), 3) Monitored EM pool trips relative to unmonitored EM pool trips (EM-EM), and 4) Monitored observer pool trips relative to all EM pool trips (OB-EM, observer data available to support EM monitoring). The OB-OB and EM-EM gap analyses were the focus of this analysis because they most closely describe whether monitored trips are representative of all trips within deployment strata. The OB-ZE and OB-EM analyses were included to assess the availability of observer pool data to other dependent pools.

Post-strata were generally defined by gear type, FMP, tender status, and the dominant species landed (trip target), with exception to the OB-EM analyses, in which tender status was excluded in the

70

post-strata definition. This was done to mimic the post-strata CAS employs to generate discard estimates for the observer, zero-selection, and EM pools (i.e., OB-OB, OB-ZE, and EM-EM) which do not necessarily match those used in average weight estimates applied to EM monitoring (i.e., OB-EM).

Within the post-strata of a given stratum, distance categories were defined for each trip as a function of whether the trip was monitored or its proximity to a monitored trip: 1) trip is monitored (MD), 2) nearest monitored trip occurs 15 days before or after the unmonitored trip in the same NMFS area (AD), 3) nearest monitored trip occurs within 45 days before or after the unmonitored trip in the same FMP (FD), or 4) the nearest monitored trip meets none of the other categories and the nearest monitored trip occurs within the same year within either FMP (YD) (Appendix Table B-1). After assigning distance categories to all trips within a given post-stratum, a single 'gap index' was calculated as a weighted proportion of trips within each of the four distance categories:

$$G_D = (P_{MD} \times 1) + (P_{AD} \times 0.75) + (P_{FD} \times 0.25) + (P_{YD} \times 0)$$

where G_D is the gap index for a given post-stratum D and P_{MD}, P_{AD}, P_{FD}, and P_{YD} are the proportions of trips in each distance category. The weights for the distance categories are arbitrary but were specified to provide separation between the AD distance category to the FD and YD categories that aids in interpreting whether or not a post-stratum has adequate coverage for generating area-level estimates. The gap index represents an overall measure of the spatiotemporal availability of monitoring data within a given post-stratum.

The realized strata-specific deployment rates in 2019 (Table 5) were used to inform simulated trip deployment. In trip simulations, *EM HAL* and *EM POT* trips were pooled to form a single deployment stratum with a selection rate of 32.5% since both of these strata rely on the same imputation and catch estimation routines. Trip simulations were performed 10,000 times (iterations) to capture the full suite of possible outcomes expected under the actual deployment rates achieved. Gap indices were calculated for each iteration to produce simulated distributions to represent the range of possible outcomes under actual 2019 trip selection rates. For a given post-stratum, the simulated distributions of gap indices were compared to the gap indices resulting from trips that were actually monitored in 2019 (the realized gap

indices). By calculating the proportion of simulated outcomes that were equal to or more extreme than the realized outcomes, post-strata with unlikely outcomes were identified.

RESULTS

Results of the gap analyses are presented in Appendix Table B-2. Summaries of key results are presented by deployment strata below. Graphic depictions of realized coverage are presented in Appendix Figure B-3 through Appendix Figure B-7 to illustrate how monitored trips were spatiotemporally distributed between post-strata in 2019 and provide context to the acquire gap indices.

Observer Monitored Hook-and-line Stratum (HAL)

The OB-OB comparison gap indices for halibut-target trips in the BSAI was on the low tail of the simulated distribution (3.2% of outcomes were at least as extreme) but the gap index for halibut-target trips in the GOA was on the high tail of the simulated distribution (8.0% of outcomes were at least as extreme) (Appendix Figure B-1). This may be due to higher realized monitoring rates in the GOA (19.6% of 152 trips) than in the BSAI (16.04% of 187 trips) and also because there was little monitoring in the Aleutian Island areas resulting in many trips being categorized in the FMP-level distance category (Appendix Figure B-3). None of the 27 trips in area 542 were observed, and only 11.29% of the 62 trips in 541 were observed (Appendix Figure B-3, Appendix Table B-2). In contrast, 28.09% of the 89 trips in area 610 were observed. The elevated observer monitoring coverage in GOA halibut-target trips also resulted in unlikely (4.87% outcomes were at least as extreme) and higher OB-EM gap indices (Appendix Figure B-1 and Appendix Figure B-6).

Pacific cod target trips in both the BSAI and GOA had gap indices in the upper ends of the simulated distributions (11.1% and 14.9% of outcomes were as high or more extreme, respectively). FMP-specific realized coverage rates were higher than the realized rate for the hook-and-line stratum as a whole (22.2% in the BSAI and 21.3% in the GOA), and further exploratory analyses indicate that the

elevated realized rates were due to a high number of inherited trips that were monitored at the beginning of the year during the Pacific cod fishery.

None of the 10 sablefish target trips in the BSAI were monitored which resulted in a gap index of zero (Appendix Figure B-3). However, this outcome was present in 14.2% of simulated iterations, so it is not wholly unexpected given fishing effort and selection rates achieved. It should be noted that these trips were generally longer in duration (mean of 13.7 days) compared to those in the GOA (mean of 4.6 days).

Observer Monitored Pot Stratum (POT – No Tender)

The realized OB-OB gap index for BSAI Pacific cod target trips was on the low end of the simulated distribution (5.2% of outcomes were at least as extreme; Appendix Figure B-2). However, the FMP-specific realized monitoring rate of 15.6% for this post-stratum was higher than the strata-specific realized monitoring rate of 14.0%, suggesting that monitoring coverage was not proportionately distributed in time and space. As evidence, none of the 12 trips in area 516 were monitored (i.e., and therefore were assigned to the FMP-level distance category) and only 1 of 39 trips in area 509 were observed in the latter half of the year that resulted in 30 of those trips being assigned to the FMP-level distance category (Appendix Figure B-4).

Observer Monitored Trawl Stratum (TRW)

The realized gap index for GOA arrowtooth-target trips was less than all gap indices from 10,000 simulations, indicating a disproportional distribution of monitoring coverage within the observer trawl stratum (Appendix Figure B-1). The realized monitoring rates for this post-stratum were only 17.2% from 233 trips, compared to the realized rate of 25.2% within the TRW stratum. GOA arrowtooth-target trips showed patterns among NMFS areas; 19.8% of the 197 trips in area 630 were monitored but only 4.8% of the 42 trips in area 620 and none of the 10 trips in area 610 were monitored (Appendix Figure B-5).

Realized gap indices for the other observer TRW post-strata were generally within simulated distributions. Of the simulated outcomes in the shallow water flatfish-target, 10.3% of simulated

outcomes were at least as extreme. This coincides with the disproportionately lower monitoring rates within arrowtooth-target trips.

Observer Monitored Trawl and Pot Tender Strata (TRW - Tender and POT – Tender)

Monitoring coverage was generally proportionately distributed across post-strata within both the *TRW – Tender* and *POT – Tender* strata (Appendix Figure B-2 and Appendix Table B-2) as indicated by realized gap indices well within the simulated distributions.

EM Monitored Hook-and-line Stratum (EM HAL)

Most of the post-strata within the EM HAL stratum within the EM-EM comparisons had realized gap indices that were on or near the tails of the simulated distributions, indicating disproportionate monitoring coverage (Appendix Figure B-1).

Trips targeting Pacific cod had realized gap indices that were on the upper ends of the simulated distributions in both FMPs; only 1.1% of simulated outcomes were at least as extreme as the realized gap index in the GOA, and similarly, 14.26% of simulated outcomes were at least as extreme as the realized gap index in the BSAI. The realized coverage rates in these post-strata were also higher than the strata-specific realized rate: 45.0% of 20 trips in the BSAI and 50.7% of 73 trips in the GOA compared to the stratum-wide rate of 32.5%.

Conversely, realized gap indices for trips targeting halibut were on the lower tails of their simulated distributions, especially in the BSAI where trips were monitored at 19.61% of 51 trips and only 3.5% of simulated outcomes had gap indices as or more extreme. Additionally, no halibut trips were monitored in the BSAI prior to late June, resulting in many trip assigned to the FMP-level distance category in areas 518, 541 and 542 (Appendix Figure B-6). Taken together, EM HAL coverage appears to have been skewed towards Pacific cod and away from halibut.

EM Monitored Pot Stratum (EM POT)

Monitoring coverage was generally uniformly distributed across post-strata within the *EM POT* strata of the EM-EM comparisons (Appendix Figure B-2 and Appendix Table B-2) as indicated by realized discard gap indices well within the simulated distributions.

DISCUSSION

This analysis indicates that the deployment of fishery monitoring tools was occasionally disproportionately distributed between post-strata. For observers this occurred within the *HAL*, *TRW- No Tender* and *POT- No Tender* strata. For EM this occurred within the *EM HAL* stratum. In this analysis low realized monitoring rates and a low gap index in the BSAI halibut-target post-stratum and high realized monitoring rates and a high realized gap index in the GOA were found. These results are consistent with the findings in the main analysis for the observer pool *HAL* stratum that found that area 542 had fewer trips observed than expected and area 610 had more observed trips than expected (Fig. 7). The low gap index in the *TRW* GOA arrowtooth-target post-stratum may have been due low monitoring rates in area 620, which coincides with results in the main section of this report where this stratum had 18 fewer trips observed than expected in the area (Fig. 9). Finally, there was a pattern within the *EM HAL* stratum in both the BSAI and the GOA where acquired gap indices for Pacific cod-target post-strata were high but acquired gap indices for halibut-target trips were low. This target-specific pattern was not apparent in the analyses in the main section of this report.

Despite these differences, this analysis also indicates that the deployment of monitoring within the observer pool generally resulted in expected overlap of observer coverage with fishing activity in the zero-selection pool and fishing activity in the EM pool. Only one post-stratum in the OB-EM analyses – *EM HAL* GOA halibut-target – had a realized gap index that was at least as extreme as 5% of simulated outcomes (Appendix Table B-2).

These conclusions only indicate that the level of coverage provided by the observer pool for the zero-selection and EM pools largely met expectations given the degree of spatiotemporal overlap between the pools and does not speak to whether or not the degree of data provided by the observer pool was or will be adequate for discard or average weight estimates. Further work is required to determine whether any findings of this analysis were present in previous years or will persist in future years. Additionally, further investigation is required to determine the specific impacts as well as the mechanisms though which any persistent patterns manifest prior to prescribing solutions.

CITATIONS

NMFS (National Marine Fisheries Service). 2019. *Draft* 2020 Annual Deployment Plan for Observers

and Electronic Monitoring in the Groundfish and Halibut Fisheries off Alaska. National Oceanic

and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available at

https://www.fisheries.noaa.gov/resource/document/draft-2020-annual-deployment-plan-observers-

and-electronic-monitoring-groundfish.

Appendix Table B-1. -- Distance categories assigned to each trip by the gap estimation routine using nearest-neighbor methods.

Category	Resolution	Condition	Weight
Monitored (MD)	Fine	Selected for monitoring	1.00
Area (AD)		<= 15 days of monitored trip within NMFS area	0.75
FMP (FD)	↓	<= 45 days of monitored trip within FMP	0.25
Year-to-Date (YD)	Coarse	> 45 days of monitored trip	0.00

Appendix Table B-2. -- Summary table of gap analyses for the observer (OB), zero selection (ZE), and electronic monitoring (EM) pools. 'Type' defines the type of coverage (e.g., OB-OB is observed OB trips relative to unobserved OB trips, and OB-ZE is observed OB trips relative to all ZE trips, etc.).'Rate' is the post-strata-specific realized monitoring rate as a percentage. Gap indices represent the spatiotemporal availability of monitoring data. 'Realized' gap indices resulted from monitored trips in 2019 and 'Min', 'Med', and 'Max' represent the minimum, median, and maximum gap indices resulting from 10,000 simulations of trip selection at realized deployment rates. 'Likli.' represents the proportion of simulated outcomes that were at least as extreme as the realized result under the assumption of random deployment. Outcomes with lower likelihood are shaded darker. SWF = Shallow Water Flatfish Target.

Type	Gear/Tender	FMP	Trip Target	Rate	Gap Indices Realized	Min	Med	Max	Likli.
OB-OB	HAL	BSAI	Halibut	16.04	0.611	0.502	0.688	0.782	0.0320
OB-OB	HAL	BSAI	Pacific Cod	22.22	0.778	0.000	0.639	0.889	0.1117
OB-OB	HAL	BSAI	Sablefish	0.00	0.000	0.000	0.355	0.921	0.1419
OB-OB	HAL	GOA	Halibut	19.74	0.787	0.723	0.773	0.801	0.0801
OB-OB	HAL	GOA	Pacific Cod	21.28	0.704	0.000	0.653	0.806	0.1496
OB-OB	HAL	GOA	Sablefish	16.78	0.778	0.729	0.776	0.803	0.4090
OB-OB	POT	BSAI	Pacific Cod	15.56	0.689	0.588	0.736	0.795	0.0516
OB-OB	POT	BSAI	Sablefish	10.81	0.586	0.000	0.599	0.829	0.4521
OB-OB	POT	GOA	Halibut	16.67	0.708	0.000	0.292	0.958	0.1183
OB-OB	POT	GOA	Pacific Cod	11.11	0.678	0.000	0.683	0.833	0.4968
OB-OB	POT	GOA	Sablefish	10.56	0.614	0.397	0.651	0.787	0.2276
OB-OB	POT - Tender	BSAI	Pacific Cod	30.77	0.745	0.133	0.765	0.893	0.3632
OB-OB	POT - Tender	GOA	Pacific Cod	16.67	0.821	0.000	0.750	1.000	0.3141
OB-OB	TRW	BSAI	Pacific Cod	29.37	0.793	0.701	0.795	0.845	0.4808
OB-OB	TRW	GOA	Arrowtooth	17.17	0.691	0.714	0.804	0.847	0.0000
OB-OB	TRW	GOA	Flathead Sole	44.44	0.600	0.000	0.500	0.975	0.2616
OB-OB	TRW	GOA	Pacific Cod	33.33	0.833	0.000	0.817	0.933	0.3059
OB-OB	TRW	GOA	Pollock	25.68	0.813	0.794	0.812	0.825	0.4733
OB-OB	TRW	GOA	SWF	35.29	0.779	0.136	0.671	0.879	0.1037
OB-OB	TRW - Tender	BSAI	Pacific Cod	50.00	0.886	0.000	0.818	1.000	0.2006
OB-OB	TRW - Tender	GOA	Pacific Cod	25.93	0.815	0.620	0.843	0.935	0.1935
OB-OB	TRW - Tender	GOA	Pollock	42.86	0.773	0.000	0.761	0.943	0.4444
OB-ZE	HAL	BSAI	Halibut	16.04	0.523	0.268	0.431	0.624	0.0972
OB-ZE	HAL	BSAI	Pacific Cod	22.22	0.286	0.000	0.293	0.443	0.4930
OB-ZE	HAL	BSAI	Sablefish	0.00	0.000	0.000	0.250	0.250	0.3741
OB-ZE	HAL	GOA	Halibut	19.74	0.718	0.591	0.715	0.746	0.4128
OB-ZE	HAL	GOA	Pacific Cod	21.28	0.675	0.000	0.675	0.750	0.5000
OB-ZE	HAL	GOA	Sablefish	16.78	0.717	0.608	0.717	0.750	0.5000
OB-ZE	POT	GOA	Pacific Cod	11.11	0.250	0.000	0.250	0.750	0.5000

Appendix Table B-2. Continued.

|									
EM-EM	HAL	BSAI	Halibut	19.61	0.574	0.309	0.707	0.875	0.0352
EM-EM	HAL	BSAI	Pacific Cod	45.00	0.862	0.000	0.812	0.925	0.1426
EM-EM	HAL	GOA	Halibut	28.20	0.801	0.758	0.813	0.846	0.1547
EM-EM	HAL	GOA	Pacific Cod	50.68	0.844	0.523	0.786	0.877	0.0114
EM-EM	HAL	GOA	Sablefish	32.15	0.804	0.763	0.814	0.849	0.2070
EM-EM	POT	BSAI	Pacific Cod	33.33	0.833	0.423	0.827	0.899	0.4400
EM-EM	POT	BSAI	Sablefish	0.00	0.000	0.000	0.625	1.000	0.4509
EM-EM	POT	GOA	Halibut	0.00	0.000	0.000	0.625	1.000	0.4437
EM-EM	POT	GOA	Pacific Cod	42.00	0.805	0.470	0.820	0.900	0.3881
EM-EM	POT	GOA	Sablefish	36.23	0.796	0.511	0.764	0.873	0.2180
OB-EM	HAL	BSAI	Halibut	16.04	0.508	0.359	0.562	0.688	0.1607
OB-EM	HAL	BSAI	Pacific Cod	22.22	0.625	0.000	0.550	0.625	0.2519
OB-EM	HAL	GOA	Halibut	19.61	0.741	0.663	0.728	0.750	0.0487
OB-EM	HAL	GOA	Pacific Cod	21.28	0.659	0.000	0.581	0.724	0.1160
OB-EM	HAL	GOA	Sablefish	16.51	0.727	0.683	0.727	0.748	0.5000
OB-EM	POT	BSAI	Pacific Cod	17.23	0.750	0.411	0.750	0.750	0.5000
OB-EM	POT	BSAI	Sablefish	10.81	0.125	0.000	0.125	0.250	0.5000
OB-EM	POT	GOA	Halibut	33.33	0.000	0.000	0.000	0.250	0.5000
OB-EM	POT	GOA	Pacific Cod	11.76	0.500	0.000	0.510	0.640	0.4950
OB-EM	POT	GOA	Sablefish	11.97	0.616	0.292	0.574	0.694	0.2180

Appendix Figure B-1. -- Acquired gap indices of hook-and-line gear (HAL; top panel) and non-tender trawl gear (TRW; bottom panel) post-strata from monitored trips in 2019 (black dashed lines) compared to gap indices resulting from 10,000 simulations of trip-selection at realized deployment rates (blue distributions, with solid blue lines representing medians). Four types of monitoring coverage are shown: OB-OB, OB-ZE, and OB-EM assessed the spatiotemporal proximity of observed trips to unobserved observer pool trips, all zero-selection pool trips, and all EM pool trips, respectively, and EM-EM assessed the spatiotemporal proximity of monitored EM pool trips to unmonitored EM pool trips.
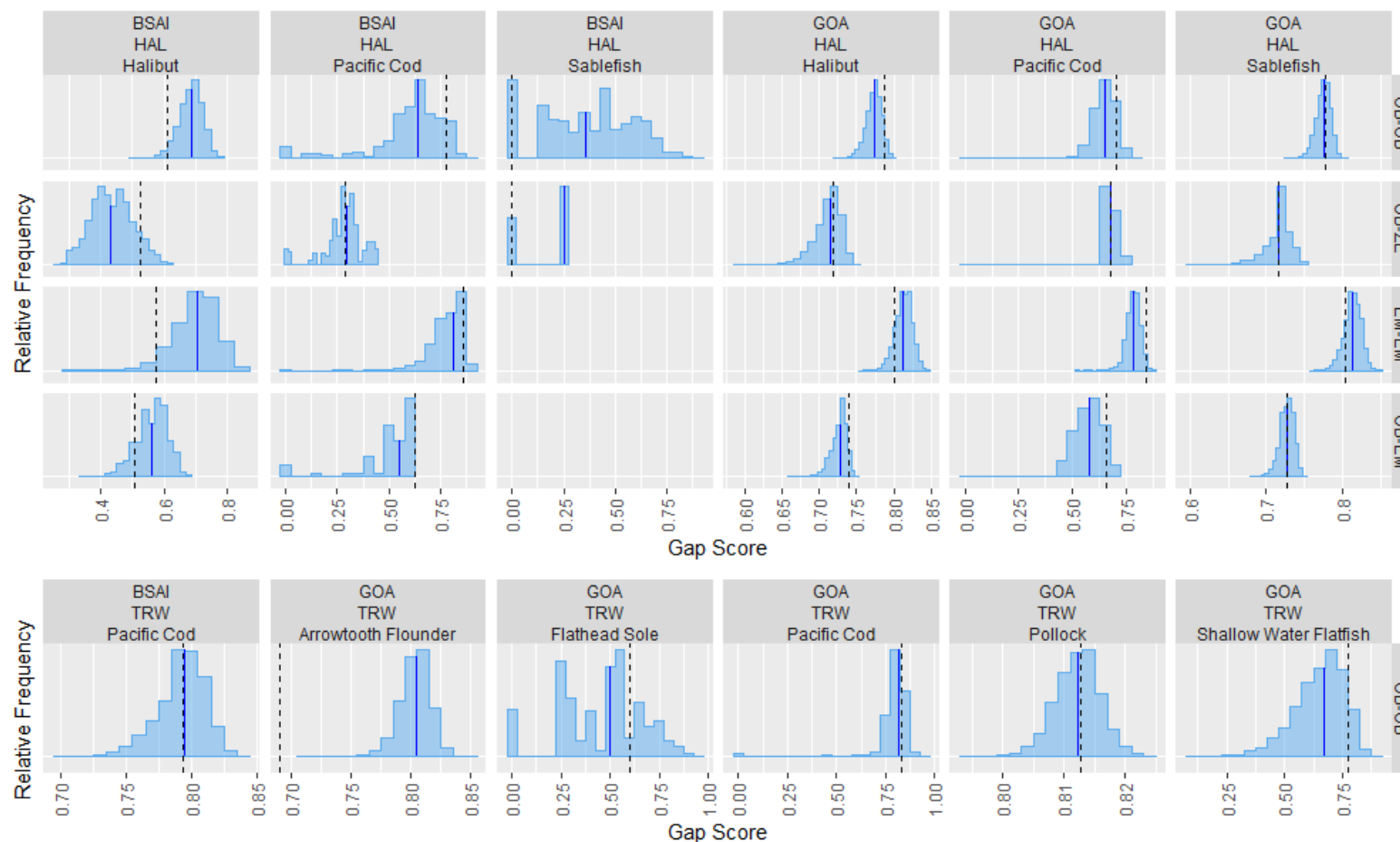
Appendix Figure B-2. -- Acquired gap indices of pot gear (POT; top panel) and tendered trawl and pot gear (TRW_TENDER, POT_TENDER; bottom panel) post-strata from monitored trips in 2019 (black dashed lines) compared to gap indices resulting from 10,000 simulations of trip-selection at realized deployment rates (blue distributions, with solid blue lines representing medians). Four types of monitoring coverage are shown: OB-OB, OB-ZE, and OB-EM assessed the spatiotemporal proximity of observed trips to unobserved observer pool trips, all zero-selection pool trips, and all EM pool trips, respectively, and EM-EM assessed the spatiotemporal proximity of monitored EM pool trips to unmonitored EM pool trips.
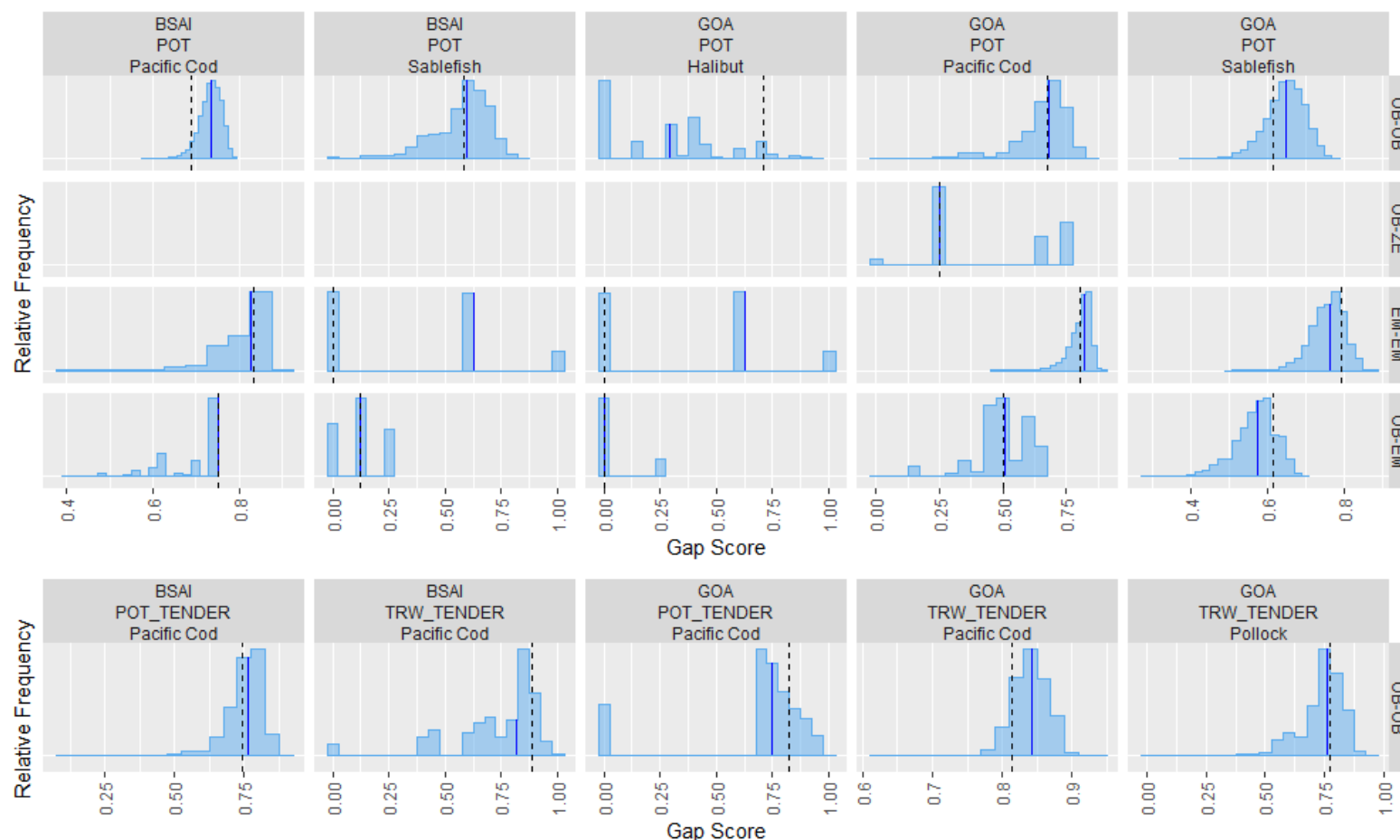
Appendix Figure B-3. -- Relative concentrations of fishing effort (red) and monitoring coverage (blue) for observer pool (OB, top) and zero-selection pool (ZE, bottom) discard gaps for 2019 hook-and-line gear (HAL) trips. Areas with fewer than 3 distinct fishing vessels were obscured and replaced with proportions of trips that were monitored.
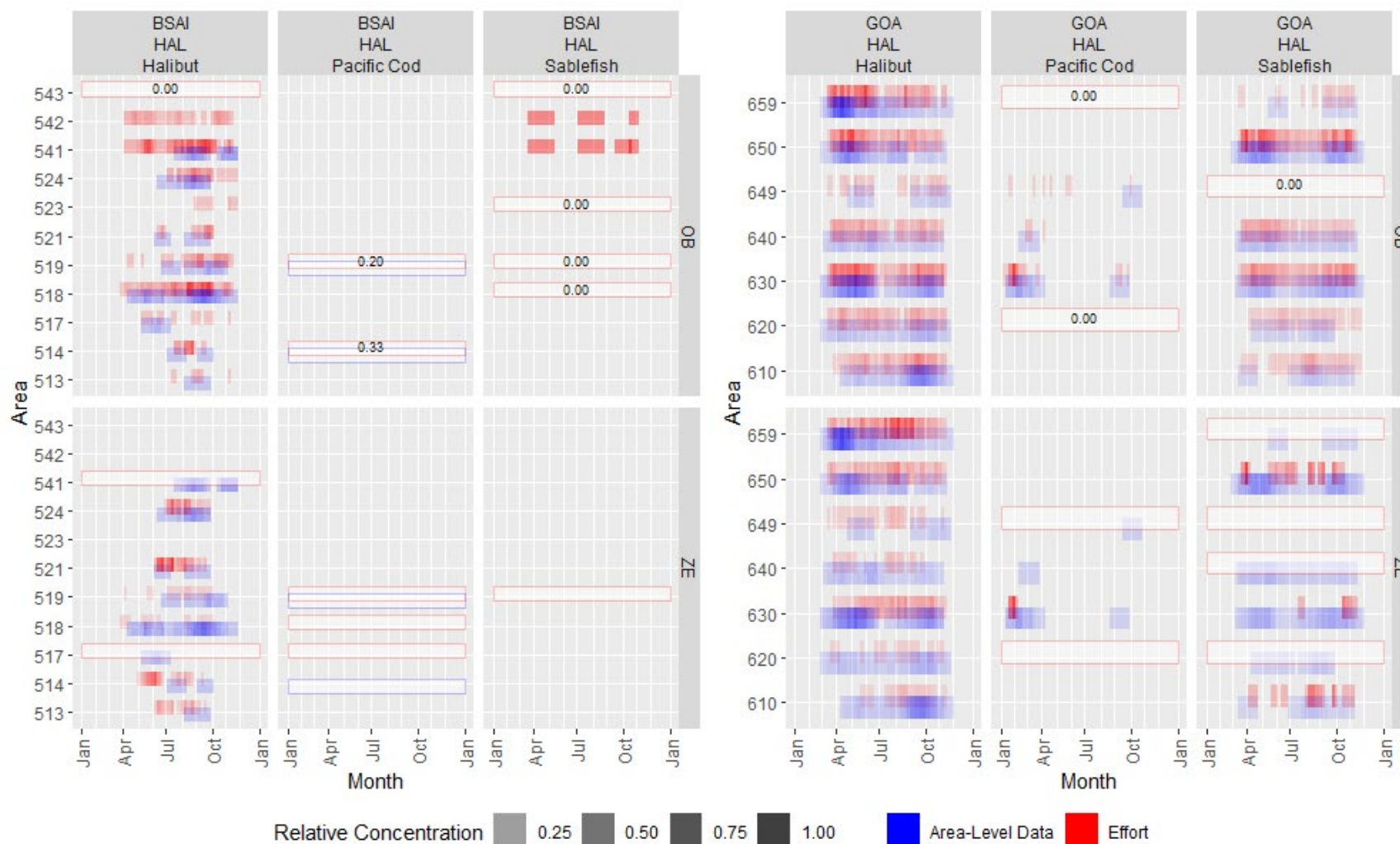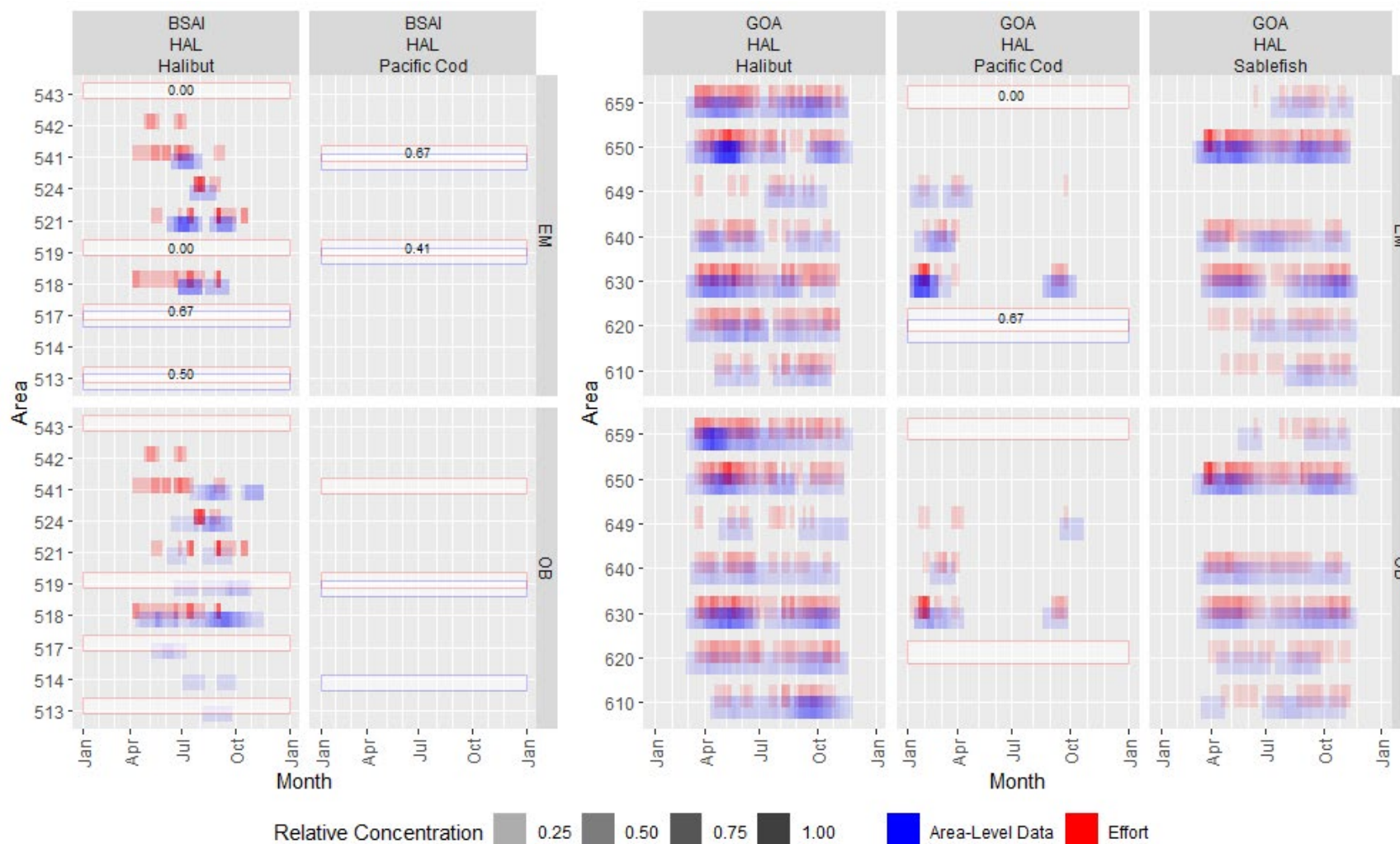
Appendix Figure B-4. -- Relative concentrations of fishing effort (red) and monitoring coverage (blue) for observer pool (OB, top) and zero-selection pool (ZE, bottom) discard gaps for 2019 pot gear (POT) trips. Areas with fewer than 3 distinct fishing vessels were obscured and replaced with proportions of trips that were monitored.
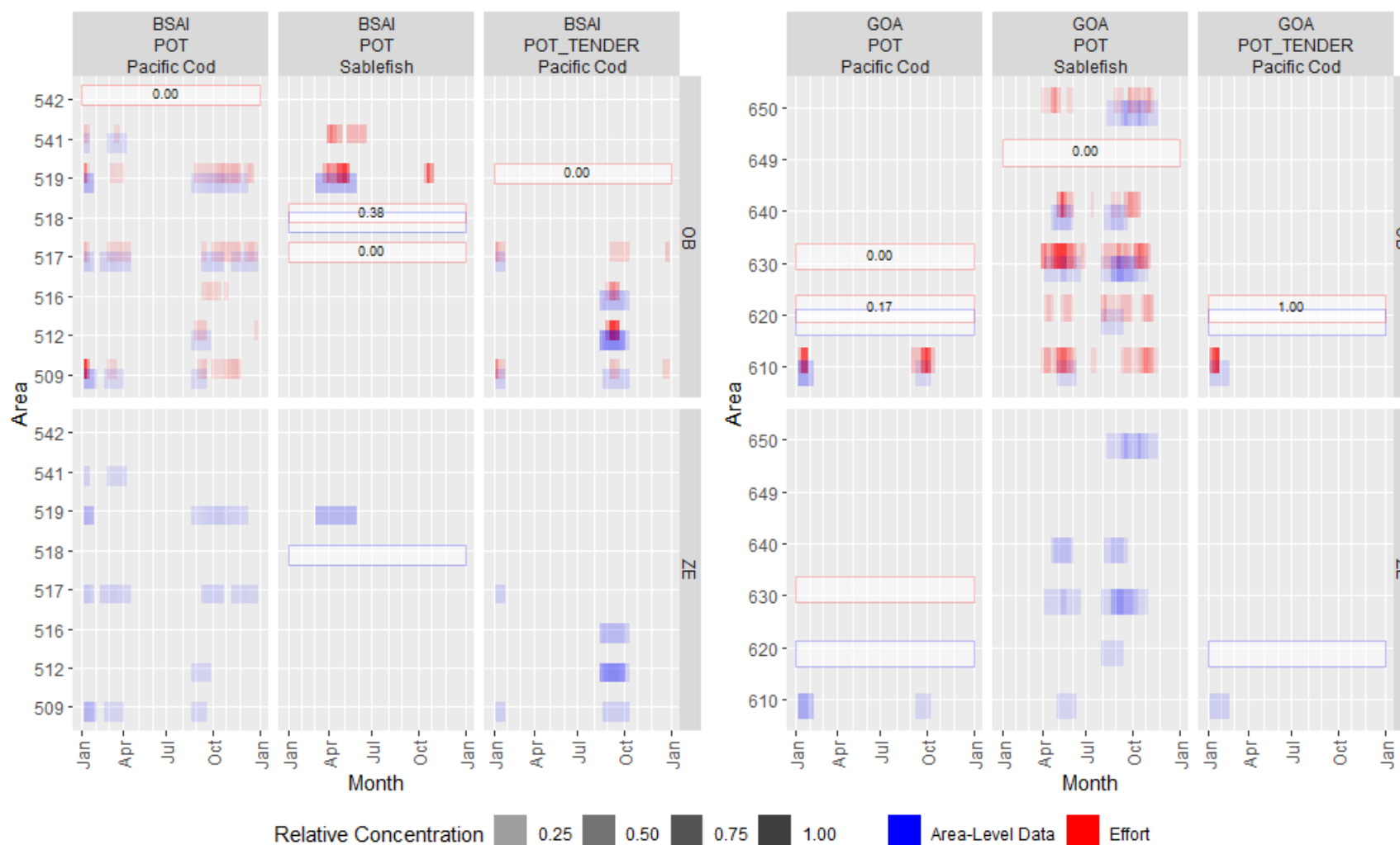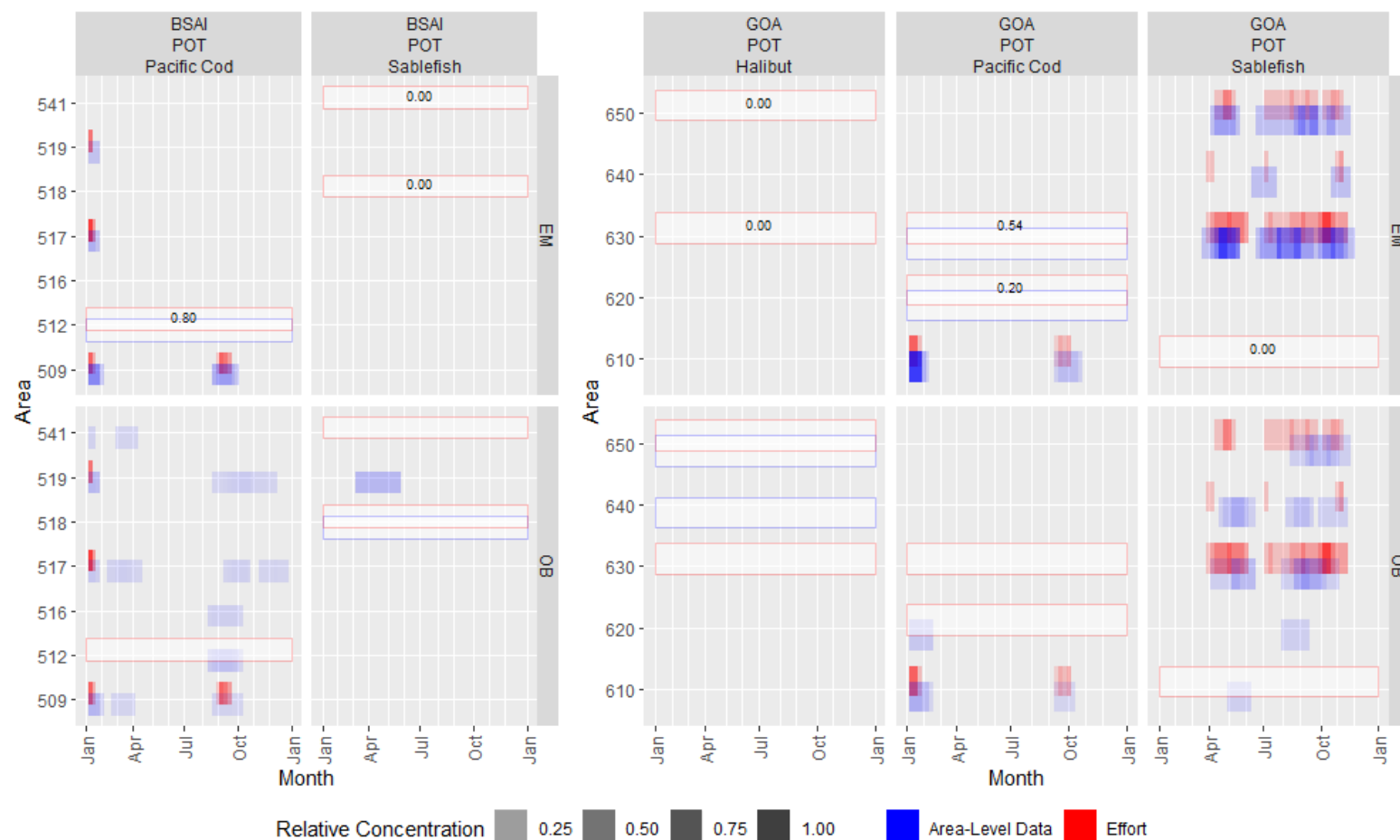
Appendix Figure B-5. -- Relative concentrations of fishing effort (red) and monitoring coverage (blue) for observer pool discard gaps for 2019 trawl gear (TRW) trips. Areas with fewer than 3 distinct fishing vessels were obscured and replaced with proportions of trips that were monitored.
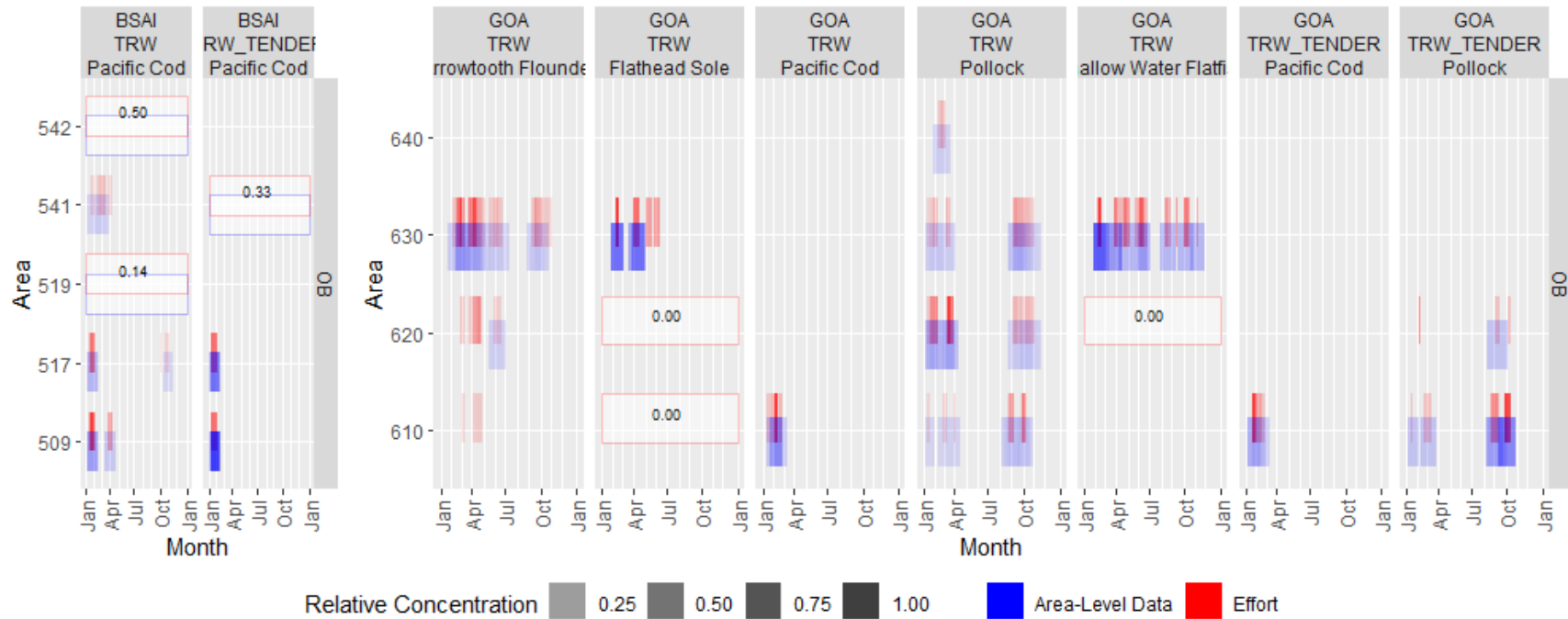
Appendix Figure B-6. -- Relative concentrations of fishing effort (red) and monitoring coverage (blue) from EM pool trips for discard gaps (EM; top) and from observer pool trips for average weight gaps (OB; bottom) for 2019 hook-and-line gear (HAL) trips. Areas with fewer than 3 distinct fishing vessels were obscured and replaced with proportions of trips that were monitored.

Appendix Figure B-7. -- Relative concentrations of fishing effort (red) and monitoring coverage (blue) from EM pool trips for discard gaps (EM; top) and from observer pool trips for average weight gaps (OB; bottom) for 2019 pot gear (POT) trips. Areas with fewer than 3 distinct fishing vessels were obscured and replaced with proportions of trips that were monitored.